165

Psychological Monographs:
General and Applied

Combining the *Applied Psychology Monographs* and the *Archives of Psychology*
with the *Psychological Monographs*

VOLUME 70
1956

Editor
HERBERT S. CONRAD
Department of Health, Education, and Welfare
Office of Education
Washington 25, D.C.

Consulting Editors

DONALD E. BAIER
FRANK A. BEACH
ROBERT G. BERNREUTER
WILLIAM A. BROWNELL
HAROLD E. BURTT
JERRY W. CARTER, JR.
CLYDE H. COOMBS
JOHN F. DASHIELL
EUGENIA HANFMANN
EDNA HEIDBREDER
HAROLD E. JONES
DONALD W. MACKINNON
LORRIN A. RIGGS
CARL R. ROGERS
SAUL ROSENZWEIG
ROSS STAGNER
PERCIVAL M. SYMONDS
JOSEPH TIFFIN
LEDYARD R TUCKER
JOSEPH ZUBIN

Published by
THE AMERICAN PSYCHOLOGICAL ASSOCIATION
1333 SIXTEENTH STREET N.W., WASHINGTON 6, D.C.

Bureau of Ednl & Psyl. Research
S. C. E. R. T.
9. 3. 81
I 737

CONTENTS OF VOLUME 70

Whole No.

408, 409, 410 A Study of Adoptive Children. I. Interviews as a Source of Scores for Children and Their Homes. II. The Predictive Validity of the Yale Developmental Examination of Infant Behavior. III. Relationships Between Some Aspects of Development and Some Aspects of Environment for Adoptive Children. J. R. Wittenborn and others. Pp. 115.

411 The Tendency to See Objects as Equidistant and Its Inverse Relation to Lateral Separation. Walter C. Gogel. Pp. 17.

412 Absolute and Relational Stimulus Training in Discrimination Learning. Sheldon J. Lachman. Pp. 21.

413 Rorschach Card Preferences as a Diagnostic Aid. Boleslaw A. Wysocki. Pp. 16.

414 The Relation Between Possession of Task-Relevant Information and Attempts to Lead. John K. Hemphill, Pauline Pepinsky, Reuben N. Shevitz, William E. Jaynes, and Charlotte A. Christner. Pp. 24.

415 Loci of Interference in a Complex Maze. Albert H. Smith. Pp. 18.

416 Studies of Independence and Conformity: I. A Minority of One Against a Unanimous Majority. Solomon E. Asch. Pp. 70.

417 Fourteen Dimensions of Temperament. J. P. Guilford and Wayne S. Zimmerman. Pp. 26.

418 Relative Visual Direction as a Factor in Relative Distance Perceptions. Walter C. Gogel. Pp. 19.

419 Transfer and Retroaction Effects of "Association Reversal" and "Familiarization" Training in Trouble Shooting. Lawrence M. Stolurow, Thomas F. Hodgson, and John Silva. Pp. 23.

420 A Content-Analysis Method for Studying Psychotherapy. Edward J. Murray. Pp. 32.

421 Sex-Role Preference in Young Children. Daniel G. Brown. Pp. 19.

422 Conditioning as Emotional Sensitization and Differentiation. Roscoe A. Dykman, W. Horsley Gantt, and John C. Whitehorn. Pp. 17.

423 Perceptual Development in a Distorted Room: A Phenomenological Study. Melvin Weiner. Pp. 38.

424 Patterns of Religious Attitude Shifts from College to Fourteen Years Later. Erland N. P. Nelson. Pp. 15.

425 Political and Religious Dogmatism: An Alternative to the Authoritarian Personality. Milton Rokeach. Pp. 43.

426 Attitudes of Young Children Toward Their Peers as Related to Certain Characteristics of Their Siblings. Helen L. Koch. Pp. 41.

427, 428 Behavioral Differences Between Normal and Traumatized Newborns: I. The Test Procedures. Frances K. Graham. II. Standardization, Reliability, and Validity. Frances K. Graham, Ruth G. Matarazzo, and Bettye M. Caldwell. Pp. 33.

429 The Sense of Effectiveness in Public Affairs. Elizabeth Douvan and Alan M. Walker. Pp. 19.

Bureau Ednl. Psy. Research
DAVID HA — AINING COLLEGE
Dated
Acc. No

No. 408
1956

Vol. 70
No. 1

A Study of Adoptive Children: I. Interviews as a Source of Scores for Children and Their Homes

By

J. R. Wittenborn and Others

Rutgers University

Bureau Ednl. Psy. Research
DAVID H... TRAINING COLLEGE
Dated ....
No ....

Vol. 70, No. 1 Whole No. 408, 1956

Psychological Monographs: General and Applied

A Study of Adoptive Children: I. Interviews as a Source of Scores for Children and Their Homes[1]

J. R. Wittenborn
Rutgers University

With the collaboration of:
Myrtle A. Astrachan, Ph.D., Psychologist
Marjorie W. DeGooyer, M.S., Psychologist
W. Wallace Grant, M.D., Pediatrician
Irma Z. Janoff, Ph.D., Psychologist
Robert B. Kugel, M.D., Pediatrician
Barbara J. Myers, M.S., Psychologist
Annaliese Riess, Ph.D., Psychologist
Ellery C. Russell, M.S., Social Worker

INTRODUCTION

THIS REPORT has its origin in a much larger investigation of adoptive children. Our major investigation was a repeat or follow-up study of children who were examined in infancy at the Yale Clinic of Child Development.[2] In general, our major study was concerned with two questions:

1. How accurately and in what respects can the Yale Developmental Examination of infant behavior (6)[3] be employed to predict future development of adoptive children?

2. What characteristics of adoptive homes (placements) may be shown to be correlated with characteristics of adoptive children?

In order to answer these questions in as dependable a manner as possible, it was decided to use two samples of adoptive children. One sample included 114 children who were about five years of age at the time of our follow-up inquiry but who had not yet become involved in the work of the first grade. The second sample included 81 children who were past the first grade but were not yet preadolescent in their development.

It is obvious that a study of Questions 1 and 2 requires the employment of three classes of information: information emerging from the infant examination, information descriptive of the

---

[1] The staff of the project wishes to express its appreciation to Dr. Donald Young, Director of the Russell Sage Foundation, and to Dr. Carl Hovland and Dr. Milton Senn of Yale University. The nature and the quality of this effort have been greatly influenced by such friends and advisers as Dr. Joseph Baldwin, Dr. Nancy Bayley, Dr. Harold Jones, Dr. Neal Miller, Dr. Robert Sears, and Dr. Helen Thompson. We are indebted also to Dr. Leonard Cottrell, Dr. Liselotte Fischer, Dr. Else Frenkel-Brunswik, Dr. Marjorie Honzik, Dr. Mary Cover Jones, Dr. Frederick Mosteller, and Dr. A. J. J. Van Woerkom. We are particularly grateful to the adoptive families who participated in this investigation, the child placement agencies who cooperated with us and guided our efforts, and to Adelaide Piscitelli, David Macklin, David Jenkins, and Peter Field who helped in the analysis of the data and the preparation of the manuscript.

[2] The Yale Clinic of Child Development (housed in the Institute of Human Relations, Yale University) is now known as the Yale Child Study Center.

[3] References for the present report are listed on page 58.

child's development at the time of our inquiry, and information descriptive of various facets of the child's environment. Measures for various facets of the infant examination procedure were available from the filed infant examination protocols. Satisfactory measures of the child's scholastic and mental development can readily be incorporated in almost any kind of follow-up study. The same may be said for certain gross aspects of physical development. Unfortunately, from the standpoint of personality development there were no convenient measuring devices for our children. Moreover, no convenient and satisfactory measuring devices were available for a broad description of the children's environment.

In this connection we emphasize the word "convenient"; rating and observational procedures are available for children, but they are too conspicuous, time-consuming, or demanding of the subject to lend themselves to the purposes of our investigation. Adoption is not casually regarded in our society, and the adoptive status of children and families is either a family secret or a tenderly regarded circumstance. As a consequence, follow-up examinations of adoptive children and parents must be conducted not only with the confidence and cooperation of the subjects but must be of such a nature as to offer no threat to the subject's peace of mind. It is obvious, therefore, that for our study repeated visits to the home, ratings in the school situation, and series of examinations were out of the question. Procedures for studying adoptive children must be particularly facile and inoffensive. The emphasis on speed and convenience was especially critical for us. Our adoptive families were widely scattered over Connecticut and, since in many cases the day available for examination was of necessity largely devoted to travel, our examination procedure had to be agreeably efficacious.

Our total approach to the investigation was such that we could arrange for two opportunities to get information descriptive of the home. One of these opportunities was the occasion of our social worker's visit to the home (at which time she indicated the nature of the cooperation we desired of the adoptive family and made provision for the mother and child to call at the Institute of Human Relations to participate in the follow-up examination for the child). Our second opportunity to acquire information came at the time of the child's examination. It was planned that a standard, semistructured interview would be conducted with the mother by one psychologist member of the staff while the psychological examinations of the child were being conducted by another staff member.

Incidentally, it was anticipated that the interview with the mother could be conducted in such a fashion as to produce much material descriptive of the child's personal development and general behavior, particularly as it is related to the home.

It appeared, therefore, that it would be possible for us to gather the kinds of information that would be essential to our investigation. Our problem then became one of how best to use such information in a convenient, dependable, and understandable form. In effect, our problem became one of *how to score an interview.*

We decided that a defensible scoring procedure would be possible if, during the course of the interview, answers to a standard set of questions were recorded. If such answers could be secured without explicitly asking the questions, presumably the answers would be little biased by the interviewee's possible awareness of our specific interests. If such answers (bits of information) could be recorded in standard form, possibly the material could be scored for several different modes of response or classes of information, e.g., in much the same fashion as personality questionnaires are scored. What was needed was a basis for scoring. It was decided that the basic scoring distinctions should emerge from the employment of a twofold criterion. Specifically, *bits of information could contribute to a given score if they were* *positively interrelated, and if they appeared to have a similar behavioral significance.* In establishing the scoring procedure, each of the various bits of information emerging from the interview could be correlated with every other one, and clusters or groups of information could be defined by combining the bits of information which were positively correlated with each other. As a second step, bits of information so grouped could be scrutinized and any bit which in its behavioral significance seemed to be inconsistent with the group could be eliminated.

In this way, bits of information were assigned to groups so that all the bits comprising a group were positively interrelated and had similar behavioral implications. As a consequence, it became a simple matter to score the interview of a given person for each of the groups or clusters. A given person's score for any given group or cluster was simply the number of component bits of information which appeared in his interview record.

As an incidental methodological point, it should be noted that negative correlations were never a basis for including a bit of information in a group or cluster, and that all bits of information that did not show a larger number of significant correlations with the other bits than could be accounted for on the basis of chance alone were automatically excluded from further consideration.

We can discuss the principle of scoring interrelated bits of information in terms of its implications for the total score. Although evidence that the component bits are consistently interrelated is an incidental evidence that the total score may be reliable, this is not the sole or even the principal value of using interrelated bits. It is possible that the principal value of scores that are built on consistently interrelated bits of information lies in the fact that the meaning of such scores is not likely to depend greatly upon the particular combination of items that happens to contribute to the total score for a particular individual. Scores based on bits of information that are known to be positively interrelated are preferable to scores based on bits of information combined on an a priori basis and without empirical evidence of internal consistency.

As an additional means of getting information relevant to the children's adjustment, an interview was planned for the children. Since more or less spontaneous and continuous discussion is often difficult with young children, no attempt was made to employ the semistructured interview with its avoidance of specific questions. Instead, a standard question-and-answer interview was planned for the children. It was hoped that such an interview (we called this a Social Reaction Interview) would provide useful information about the children's values and opinions of their own roles and reactions. The data from such an interview emerge in the form of numerous bits of information and lend themselves very well to a cluster-scoring procedure.

In the following sections the procedure for forming scores on the basis of interrelated bits will be applied in four different respects. In each instance the major application will be with our younger sample of children, but the results will be described also for our older sample. The respects in which we apply the cluster scoring principle may be briefly listed:

1. *Scores based on the Social Reaction Interview.* This interview was rigidly

structured and comprised a series of "What would you do if—" questions which a psychologist asked each child at the time of our follow-up study.

2. *Scores descriptive of the child based on the mother's Child-Rearing Practices Interview.* In our Child-Rearing Practices Interview with the mother a number of areas were introduced for discussion, but specific questions were generally avoided. Nevertheless, the interviewer sought and recorded specific bits of information descriptive of the child. These bits of information provided scores which described the child as seen by the mother, a very important descriptive feature of the adoptive child.

3. *Scores descriptive of the environment based on the mother's Child-Rearing Practices Interview.* Much of the psychologist's interview with the mother was descriptive of the psychological environment which the home provided for the child. This information, although not sought by explicit questioning, was recorded in standard form and provided the basis for a cluster-scoring procedure.

4. *Other scores descriptive of the home (Conditions of Child Rearing).* As a preparation for the visit of the mother and the child to the Institute of Human Relations, our social worker visited the home and on the basis of this visit gathered numerous bits of information in standard form. At the time the mother brought the child in for follow-up examination, the values and the organization of the home were the subject of discussion with the psychologist and additional bits of information were gathered in standard form. This body of information, emerging both from the social worker's visit and from the psychologist's interview, provided the basis for additional cluster scores which were primarily descriptive of the home, its organization and values.

SCORES BASED ON THE SOCIAL REACTION INTERVIEW

In Appendix A will be found a series of questions which ask the child what he would do in certain standard situations, for the most part situations familiar to him. The procedure was one wherein the question was asked by the examiner and the child's response was classified under the alternatives which had been set up during our preliminary pretesting work. The children responded to this series of questions with promptness and pleasure. It cannot be claimed that children's replies describe their behavior, but sensible and socially acceptable replies by a child indicate that he knows what society expects of him and that he is willing to say this in the context of the interview. There are numerous alternative responses to each question; some of them may have a wide and general acceptance as proper in our society. Presumably this procedure would reveal great differences, if they exist, in the children's system of values or ideas concerning appropriate behavior. Such verbalizations are not irrelevant. For example, most of the behavior difficulties that adults bring into the clinic and much of the psychotherapeutic procedure that is employed there concern how an individual feels he would or should react in common situations.

The Social Reaction Interviews for 100 children from our younger sample were reviewed and it was found that 111 response items occurred with sufficient frequency to justify further scrutiny. Accordingly, each of these items was correlated with every other one and the intercorrelations were examined for clustering tendencies. As a result of such examination, several clusters of items with common behavioral implications were discernible. For the most part, items that have generally high correlations were excluded because of the difficulty of assigning them to any particular cluster; and items which are correlated with only one or two other items were excluded, because their relevance to any one cluster was not sufficiently general. (In addition, the reliability of such items, as inferred from the correlations, was questioned.) It should be emphasized that after the groupings of items were formed on the empirical basis of positive intercorrelations, the items were scrutinized from a rational standpoint and the intercorrelated items that did not plausibly go together (i.e., which could not be conceptualized as being evidence of a common form of behavior) were excluded from the clusters. Eventually it was possible to organize groups of items that were positively interrelated with each other, not importantly related with numerous items in another group, and that also had a plausible, common core of meaning or inferable behavioral relevance.

After such clusters were established for the younger sample, it was possible to score the Social Reaction questions for all of our children on the basis of each of these clusters. Such scores then would express the degree to which the kind of behavior implied by a cluster was characteristic of a child. Perhaps it is better to describe these scores as extensive scores rather than intensive scores. They do not imply how intense or severe any one response of a child may be, but instead indicate how general or pervasive a given kind of response is for each child.

Tables 1 through 6 are descriptive of each of the clusters which were derived from the Social Reaction Interview with the younger children. These tables indicate the general nature of the items and also the degree to which the items in the various clusters are intercorrelated with each other. The interested reader may examine the meaning of the items more fully by referring to the Social Reaction questions in Appendix A. The intercorrelations among the various scales or cluster scores derived from the Social Reaction Interview are presented in Table 7.

The reader may question the nature of the cluster scores. The meaning of the cluster scores is an inference based upon the nature of the items which comprise the cluster. Perhaps each of these clusters may have a slightly different meaning for different readers. Although we attempt to verbalize our conceptualizations of these clusters in the paragraphs to follow, the reader is free to conceptualize them differently if he so desires. This feature is not unique to our cluster scores. People conceptualize the meaning of intelligence test scores differently, too. It should be noted that all of these intercorrelations are based upon a relatively small sample and are, of course, subject to the kinds of sampling fluctuation that ordinarily characterize such data. Accordingly, the reader should not hasten to ascribe some unique significance to the highest correlation, nor to feel that a zero or slightly negative correlation is necessarily contrary to the general implications of the positive interrelationships.

It may be suspected that we have built our clusters adventitiously on the basis of correlations that emerged largely as the result of sampling factors, and that one should be most hesitant in accepting intercorrelations among these items as an evidence for their behavioral similarity. Actually, however, all the items that appear in our cluster scores showed many more large correlations with the other items in the Social Reaction Interview than reasonably could be anticipated on the basis of chance alone. As a matter of fact, the appearance of consistently interrelated bits of information is important evidence that the bits so interrelated are not combined on the basis of chance. Consistency and randomness simply do not go together.

Some of the clusters are of sufficient intrinsic interest to justify a few words of discussion. For example (see Table 1), *Dependence on Adults* is remarkable in that all of the items describe a similar type of response, although the various situations that motivate or cue the responses are seemingly different from each other. The consistency of this form of behavior or attitude is indicated by the positive intercorrelations among the items.

The cluster in Table 2, which we have called *Aggression,* comprises ten items which may be seen by many readers to have aggressive implications. It is quite possible, however, that other readers will be impressed with the general implication of strong will or insistence on carrying out one's own plan that is inherent in most of these responses. It seems in general, however, that more than stubbornness or strong will is implied by these items. For example, many of them imply a readiness to employ action that will be contrary to the wishes of others and possibly punitive in nature. In Table 2 we find a *question mark* in one of the cells; question marks will be found in several of the tables. These question marks are due to the fact that in some instances the marginal frequencies in our fourfold tables were of such a nature as to make the determination of the tetrachoric correlation coefficients quite uncertain (we used computing diagrams). Accordingly, we include a question mark in our table.

Although there are relatively high intercorrelations among the items comprising the cluster shown in Table 3, the term *Socialized Compliance* may not be equally agreeable to all readers. Some may see neatness in personal relations or avoidance of differences with others as the underlying component in these responses. Still others may be inclined to say that the responses merely indicate a goody-good pattern of verbalization. Although these alternative interpretations may have merit, it seems that no great stretch of imagination is required to describe these various responses as involving compliance, on the verbal level at least, with some of the polite expectations for children in middle-class Western society.

The intercorrelations for the items comprising Table 4 are not uniformly high, but on the whole they indicate that there is some empirical similarity among them. It seemed to us, at least, that from a rational standpoint these items have something in common that involves doing the kind of thing that responsible adults might do in a child's view of adults. The responses imply a certain quality of concern with others which many fortunate children have known in the behavior of adults. We call this cluster *Taking an Adult Role.*

We have designated the cluster in Table 5 as *Weakness-Avoidance.* It is not very good from the empirical standpoint; some of the interrelationships could not be quantified because the marginal frequencies made the determination of the tetrachoric coefficients quite uncertain. In addition, there are several zero correlations. Nevertheless, it seems desirable to us to take this opportunity to form a score which might indicate weak, avoidant, hopeless, or futile attitudes on the part of a child.

The cluster in Table 6 provides a basis for a *Constructive Approach* variable. In its apparent implications it is similar to and empirically highly correlated with some of the other clusters, particularly the *Taking an Adult Role* cluster.

The intercorrelations among these various cluster scores are shown in Table 7. It is quite evident from Table 7 that, although they involve few overlapping items, these cluster scores are far from independent of each other. The most conspicuous evidence of relationship is the one between *Constructive Approach* and *Taking an Adult Role.* It appears that *Socialized Compliance, Constructive Approach,* and *Taking an Adult Role* are all fairly well intercorrelated. It is quite interesting to note that there is an inverse relationship between *Dependence on Adults* and *Aggression,* and that the correlations imply that

TABLE 1

TETRACHORIC INTERCORRELATIONS OF RESPONSE ITEMS IN THE CLUSTER, "DEPENDENCE ON ADULT"

(Social Reaction Interview, Younger Sample; $N = 100$)

Response Number	Response Item	% Frequency of Response	Response Number																
			6d	8c	1c	2e	6b	7b	8b	9d	1d	3h	4d	6d	8b	2c	3b	1b	2b
Dependency																			
6d	Tell adult if toy breaks	24	—																
8c	Tell adult if he hurts himself	45	.63	—															
Aggression																			
1c	Ask if he wants friend's toy	34	.07	.32	—														
2e	Tell adult if child won't let him play with his toys	11	.35	.29	.20	—													
6b	Tell adult if child hits him	38	.53	.32	.14	.50	—												
7b	Tell adult if child takes his toy	40	.50	.47	.43	.72	.66	—											
8b	Tell adult if child calls him names	40	.50	.25	.10	.34	.71	.76	—										
Submission																			
9d	Tell adult if children won't let him play with them	17	.23	.13	.21	.80	.35	.70	.50	—									
Sympathy																			
1d	Tell adult if child hurts himself	33	.55	.29	.13	.32	.68	.58	.52	.25	—								
3h	Tell adult if children hit another child	30	.40	.25	.42	.50	.47	.67	.55	.48	.70	—							
4d	Tell adult if small child is crying	30	.34	.18	.30	.24	.50	.50	.50	.52	.70	.75	—						
6d	Tell adult if child breaks his toy	18	.48	.19	.10	.34	.40	.45	.52	.48	.48	.70	.62	—					
8b	Tell adult if bird is hurt	15	.30	.45	.10	.67	.35	.62	.51	.47	.53	.40	.40	.32	—				
Cleanliness																			
2c	Tell adult if ice cream gets on clothes	19	.45	.40	.41	.57	.25	.58	.49	.64	.36	.50	.40	.62	.53	—			
3b	Tell adult if ball falls in garbage can	13	.60	.13	.09	.12	.37	.20	.20	.43	.32	.27	.40	.40	.20	.37	—		
Sex																			
1b	Tell adult if girl pulls up dress	22	.45	.00	.12	.25	.46	.65	.50	.38	.33	.62	.62	.54	.35	.41	.63	—	
2b	Tell adult if boy pulls down pants	29	.43	.00	−.07	.51	.57	.75	.52	.51	.41	.55	.55	.55	.31	.16	.50	.92	—
Toilet																			
1c	Tell adult if child wets his pants	44	.50	.27	.34	.54	.52	.68	.52	.34	.56	.70	.60	.82	.26	.50	.50	.86	.89

TABLE 2

TETRACHORIC INTERCORRELATIONS OF RESPONSE ITEMS IN THE CLUSTER, "AGGRESSION"

(Social Reaction Interview, Younger Sample; $N=100$)

Response Number	Response Item	% Frequency of Response	Response Number								
			3e	4e	5f	6a	7c	8a	8c	7f	8g
Aggression											
3e	Go anyway if father breaks promise to take him to circus	16	—								
4e	Get ice cream anyway if mother says he can't have it	18	.76	—							
5f	Tell child to leave if everyone plays with child and ignores him	12	.20	.15	—						
6a	Hit child back if he hits him	37	.50	.33	.20	—					
7c	Take back his toy if another child takes it	24	.26	.30	.17	.26	—				
8a	Hit child who calls him a bad name	14	.13	.47	.40	.42	.22	—			
8c	Say something bad to child who calls him a bad name	22	.20	−.12	.32	.40	.37	?*	—		
Submission											
7f	Wouldn't go last if friends said his turn was last	11	.24	.20	.09	.11	.05	.10	.25	—	
8g	Get another friend if mother objected to one he likes	13	.41	.40	.10	.39	−.20	.38	.00	−.10	—
Sympathy											
3f	Hit children if they hit another child	17	.04	−.01	.00	.38	.22	.25	.39	.03	.29

* A "?" signifies that the frequencies of the two responses were such as to make the estimation of their coefficient quite uncertain.

Taking an Adult Role involves an aggresive attitude. Although these intercorrelations and the items that provide a basis for them are of intrinsic interest and will justify further discussions in some later report, perhaps little more need be said about them at this time.

When confronted with a new scoring device or measuring procedure, one must not permit the interest of the content to

TABLE 3

TETRACHORIC INTERCORRELATIONS OF RESPONSE ITEMS IN THE CLUSTER, "SOCIALIZED COMPLIANCE"

(Social Reaction Interview, Younger Sample; $N=100$)

Response Number	Response Item	% Frequency of Response	Response Number								
			2g	5a	1b	5b	7a	8c	9c	1f	7h
Aggression											
2g	Play with his own toys if other child won't let him play with his	26	—								
5a	Play alone if everyone plays with other child and not with him	12	.50	—							
Submission											
1b	Eat food he doesn't like	35	.00	.35	—						
5b	Give toy to other child if mother tells him to	47	.12	.42	.60	—					
7a	Go last if friends say his turn is last	31	.25	.41	.56	.63	—				
8c	Wouldn't play with child if mother tells him not to	37	.20	.32	.41	.54	.46	—			
9c	Play with someone else if other children won't let him play	17	.40	.32	.30	.30	.30	.37	—		
Sympathy											
1f	Help child who has hurt himself	40	.20	.15	.34	.40	.26	.35	.22	—	
7h	Assist or be quiet if mother is sick	31	.24	.04	.09	.50	.10	.34	.29	.40	—
Miscellaneous											
4i	Feel grown up if he were suddenly grown up	12	.12	.12	.59	.55	.03	.43	.32	.40	.63

TABLE 4

TETRACHORIC INTERCORRELATIONS OF RESPONSE ITEMS IN THE CLUSTER, "TAKING AN ADULT ROLE"

(Social Reaction Interview, Younger Sample; $N = 100$)

Response Number	Response Item	% Frequency of Response	Response Number								
			8c	1f	3f	4e	6g	2b	1a	2a	1b
Aggression											
8c	Say something bad to child who calls him a bad name	22									
Sympathy											
1f	Help child who has hurt himself	40	.44	—							
3f	Hit children who are hitting another child	17	.39	.60	—						
4e	Help child who is crying	25	?[*]	.52	.42	—					
6g	Fix toy or buy another if child breaks his toy	20	.19	.36	.33	.10	—				
Cleanliness											
2b	Clean ice cream off clothes	56	.15	.24	.26	.16	.65	—			
Sex											
1a	Tell girl to pull down dress	30	.47	.30	−.13	.30	.20	−.06	—		
2a	Tell boy to pull up pants	22	.23	.36	.04	.06	.28	−.10	.92	—	
Ethics											
1b	Replace or fix something he breaks belonging to mother	27	.00	.25	.16	.03	.60	.52	.24	.20	—
Miscellaneous											
4i	Feel grown up if he were suddenly grown up	12	.35	.40	.33	.30	.40	.60	.31	.06	.37

[*] A "?" signifies that the frequencies of the two responses were such as to make the estimation of their coefficient quite uncertain.

TABLE 5

TETRACHORIC INTERCORRELATIONS OF RESPONSE ITEMS IN THE CLUSTER, "WEAKNESS-AVOIDANCE"

(Social Reaction Interview, Younger Sample; $N = 100$)

Response Number	Response Item	% Frequency of Response	Response Number								
			3b	4g	5b	2f	6d	2c	3b	1f	2g
Aggression											
3b	Cry if father breaks promise to take him to circus	17	—								
4g	Cry if mother says he can't have ice cream	13	.80	—							
5b	Cry if everyone plays with another child and ignores him	12	.84	.78	—						
Sympathy											
2f	Avoid watching child get spanked	19	?[*]	−.09	?	—					
6d	Tell an adult if child breaks his toy	18	.14	.27	.30	.20	—				
Cleanliness											
2c	Tell an adult if ice cream gets on clothes	19	.11	.23	.28	.05	.62	—			
3b	Tell an adult if ball falls in garbage can	13	.54	.40	.29	.10	.40	.37	—		
Sex											
1f	Avoid little girl pulling up dress	10	?	?	−.05	.74	.04	−.20	−.09	—	
2g	Avoid little boy pulling down pants	14	?	?	?	.55	.08	−.10	−.20	.90	—
Miscellaneous											
1b	Believe children hate animals and insects most	9	.30	.20	.24	.07	−.16	.41	.40	.04	−.07

[*] A "?" signifies that the frequencies of the two responses were such as to make the estimation of their coefficient quite uncertain.

TABLE 6

TETRACHORIC INTERCORRELATIONS OF RESPONSE ITEMS IN THE CLUSTER, "CONSTRUCTIVE APPROACH"

(Social Reaction Interview, Younger Sample; $N = 100$)

Response Number	Response Item	% Frequency of Response	Response Number 6c	7d	8d	2g	8g	9c	6g	2b	1b (Nudity)
Dependency											
6c	Get another toy if his toy breaks	18	—								
7d	Look for another friend if his friend leaves	21	.27	—							
8d	Look after himself if he is hurt	20	.18	.21	—						
Aggression											
2g	Play with his own toys if child won't let him play with his	26	.00	.43	−.03	—					
Submission											
8g	Get another friend if mother says he can't play with someone he likes	13	.50	.44	.48	.10	—				
9c	Play with someone else if other children won't let him play with them	17	.14	.60	.31	.40	.53	—			
Sympathy											
6g	Fix toy or buy another if child breaks his toy	20	.29	.50	.00	.09	.47	.20	—		
Cleanliness											
2b	Clean ice cream off his clothes himself	56	.28	.20	.25	.35	.63	.15	.65	—	
Nudity											
1b	Ask mother later if she is taking a bath	23	.23	.33	.20	.20	.15	.01	.36	.18	—
Ethics											
1b	Replace or fix something he breaks belonging to mother	27	.50	.50	−.05	.35	?*	.46	.60	.52	.35

* A "?" signifies that the frequency of the two responses were such as to make the estimation of their coefficient quite uncertain.

obscure the practical limitations that invariably accompany limited reliability. The most desirable evidence of reliability would be provided by an arrangement where each child was given the same interview by two different examiners. This is difficult to provide under usual circumstances, and under the circumstances of our study it was impossible. Accordingly, we must be guided by estimates and indirect indications of possible reliability. As we have indicated,

TABLE 7

PRODUCT-MOMENT INTERCORRELATIONS AMONG CLUSTER SCORES
(Social Reaction Interview, Younger Sample; $N = 114$)

Cluster	Split-half Reliability*	Cluster 1	2	3	4	5
1. Dependence on Adults	.88	—				
2. Aggression	.58	−.41	—			
3. Socialized Compliance	.70	.08	.14	—		
4. Taking an Adult Role	.69	−.18	.45	.53	—	
5. Weakness-Avoidance	.57	.35	−.31	−.07	−.33	—
6. Constructive Approach	.58	−.01	.31	.46	.63	−.23

* Based on an odd-even correlation corrected by the Spearman-Brown formula.

consistent intercorrelations among the component bits comprising the cluster score are an indication of reliability, and the principle of internal consistency can be concisely applied by determining a split-half reliability by the Spearman-Brown procedure. This, however, may give an estimate which is too low if the number of items is few and if the items are somewhat heterogeneous in nature. Nevertheless, we have determined such split-half reliabilities, and they are included in Table 7, which gives the intercorrelations among the various scores emerging from the Social Reaction procedure. High correlations with other independent scores can also be indications of reliability. (It will be noted, for example, that the Constructive Approach cluster score has a correlation which exceeds our estimate for its split-half reliability.) In general, it is apparent that some of these cluster scores may have a reliability that compares favorably with many of the devices in common use for the evaluation of children. Other cluster scores offer modest reassurance of their reliability, but even for them the indications of reliability are sufficiently favorable to justify the use of the cluster scores in exploratory investigations where one hopes to get some indication of the nature and the direction of relationship, but does not hope for a close approximation of the actual degree of relationship.

The questions comprising the Social Reaction Interview for the younger children and for the older children were somewhat different from each other; a principal difference was the addition of new questions to the Social Reaction Interview for the older children. Another source of difference for the Social Reaction Interview clusters was the change in frequency of certain responses which seemingly accompanies change in age.

Changes due to age are well illustrated by a consideration of the *Dependence on Adult* cluster which was so conspicuous in our sample of younger children. In our younger group this cluster consisted of 18 items. In our older group, children mostly of eight or nine years of age, the frequency of these respective responses was so reduced that only four were sufficiently prevalent to justify their inclusion in the survey of the intercorrelations among the items of the Social Reaction Interview. When these four items were intercorrelated, they were found to bear minor positive relationships with each other. Although the implications of the *Dependence on Adult* cluster are somewhat similar for the two age groups, dependency on adults, as an attitude or point of view, obviously has less general descriptive significance for older children than for younger children. This is not surprising.

In our study of the Social Reaction Interview responses of five-year-old children, it was found that ten responses descriptive of rather simple *Aggression* were consistently interrelated and provided a basis for a cluster. These simple responses were relatively infrequent among the older children, however, and only one of them occurred with sufficient frequency to justify inclusion in our intercorrelations for the Social Reaction Interview. Since aggression in its simplest form is generally taboo in our society, it is not surprising that older and better socialized children would be reluctant to offer simple aggressive responses in the context of the Social Reaction Interview. Perhaps an additional factor which tends to suppress such responses is the kind of disciplinary control which is a part of the usual classroom situation, i.e., the classroom is possibly the earliest *generally* experienced situation wherein direct physical aggression among members of a group is uniformly tabooed and that taboo is enforced.

The trends for the *Socialized Compliance* cluster provide an interesting contrast with the trends discussed in the preceding paragraphs. The items that provide the basis for the *Socialized Compliance* cluster found for the younger children are more prevalent among the older sample than among the younger sample, and they tend to retain the pattern of positive interrelationships found for the younger children. The general tendency for a greater prevalence of the items of this cluster to be found among the responses for the older children is probably an expression of the specific learning and general socialization that tends to characterize children as they increase in age; it may also be seen as a part of the general social developmental pattern which we have previously found expressed as a reduction of aggressive responses among the older children.

In our younger sample we found a cluster of ten items that we interpreted as an assumption of an *Adult Role* in the children's attitudes. In general, it appears that the intercorrelations among the items in question are low and inconsistent for the older sample (nevertheless, the suggested pattern of relationships within the cluster appears to be similar to the pattern for the younger sample).

In the younger sample ten items were found to be interrelated with sufficient consistency to result in our combining them into a cluster which we call *Weakness-Avoidance*. The responses that comprise this cluster are quite infrequent among our sample of older children, and only one of them was sufficiently prevalent to be included in our analysis for the older group.

A final cluster comprising ten items was found to be consistently interrelated in the responses from the younger children. The nature of these items was such as to result in our calling the cluster *Constructive Approach*. These items were more prevalent among the older children than among the group of younger children, but the correlations among these items were not consistent for the older sample.

The foregoing paragraphs provide a discussion of the manner in which the organization of social attitudes characteristic of younger children may have changed for older children. It is evident from our discussion that some patterns of reaction are of no descriptive value for older children because the component behavior is relatively infrequent. Other patterns of reaction descriptive of younger children lose their descriptive value because in the older children behavior may be organized differently than it is in the younger children.

Concepts that have descriptive value for one age may provide no descriptive economy for some other age. We find that in a study of responses of older children new concepts emerge with a descriptive value that may not be applicable to younger children. In Table 8, for example, we see a group of intercorrelations for seven responses which are consistently interrelated and imply that some children are generally willing to give *"Goody-Goody"* responses to questions in the Social Reaction Interview. These very nice and obliging responses should not be confused with the overt behavior of the children in the respective situations. It should be noted that three of these responses are descriptive of school situations, but the formation of this cluster may not be due entirely to the availability of new types of items for older children. Five of the items were available for the younger children, but the responses of the younger children were not systematically related in the respects implied by these responses from the older children. Thus there appears a new pattern of response that could not have been anticipated from the responses of younger children. The *Goody-Goody* cluster is reminiscent of the *Socialized Compliance* cluster and probably has a similar general meaning. A possible difference between the two would emphasize the compliant features of the *Socialized Compliance* cluster for the five-year-olds and the "plausible" features of the *Goody-Goody* cluster for the older children.

Eight responses, considered descriptive of a generally *Responsible Attitude,* were positively intercorrelated for the sample of older children (Table 9). This *Responsible Attitude* cluster for the older children has an implication quite similar to the *Constructive Approach* cluster for the younger children. Despite their common implications, these clusters for the younger sample and for the older sample have but three common items. If one were inclined to infer that these two clusters were descriptive of the same kind of drive in the two samples, i.e., a drive to react constructively to the values and needs of one's associates, we would say also that the behavior appropriate for the expression of this drive is different for children who differ in age. Although this interpretation is plausible enough, the degree of overlap between the *Constructive Approach* cluster for the younger children and the *Responsible Attitude* cluster for the older children is not sufficient for us to be confident that a common drive pattern is implicit. We can only say that although the two sets of responses are different from each other, they appear to have a highly similar social role.

Ten of the items for the older sample are sufficiently interrelated statistically and sufficiently consistent in their meaning to result in our setting them aside as a *Cooperation with Authority* cluster (Table 10). Four of these items appeared in the *Socialized Compliance* cluster which was found for the younger children. The content of two of the items is unique to the Social Reaction Interview procedure for the older sample and could not have appeared in

TABLE 8

TETRACHORIC INTERCORRELATIONS OF RESPONSE ITEMS IN THE CLUSTER "GOODY-GOODY"

(Social Reaction Interview,* Older Sample; $N=50$)

Response Number	Response Item	% Frequency of Response	Response Number 2f	1a	2a	3b	3a	6c
Sympathy								
2f	Avoid watching child get spanked	22	—					
Sex								
1a	Tell girl to pull down dress	41	.34	—				
2a	Tell boy to pull up pants	41	.34	.85	—			
3b	Say "no" if man offers a present to go in his car	82	.29	−.05	−.15	—		
School								
3a	Explain circumstances if he got to school late	49	.06	.25	.25	.36	—	
6c	Be quiet if teacher wants everyone to be quiet	56	.24	.20	.33	.38	.10	—
Miscellaneous								
5a, b	Mother or father is person he thinks he is most like	27	.30	.30	.30	.23	−.03	.37

---

* The Social Reaction form included in the Appendix is the one used with the younger children. The form employed with the older children was identical except that it included a few additional items.

TABLE 9

TETRACHORIC INTERCORRELATIONS OF RESPONSE ITEMS IN THE CLUSTER, "RESPONSIBLE ATTITUDE"

(Social Reaction Interview,* Older Sample; $N=50$)

Response Number	Response Item	% Frequency of Response	Response Number 1f	6g	2b	3a	1b	4e	9a
Sympathy									
1f	Help child who has hurt himself	74							
6g	Fix toy or buy another if child breaks his toy	58	.50	—					
Cleanliness									
2b	Clean ice cream off clothes	76	.34	.23	—				
3a	Get ball out of garbage can if it falls in	61	.15	.04	.30	—			
Ethics									
1b	Replace or fix something he breaks belonging to mother	54	.30	.24	.40	.45	—		
4e	Replace broken window if broken while playing ball	58	.45	.08	.00	.43	.64	—	
School									
9a	Pick up coat if someone knocks it on floor	50	.38	.20	.00	−.16	.31	.25	—
Assertion									
2a	Would volunteer if teacher asked someone to lead the class	47	.02	.30	.10	.22	−.09	.12	.27

---

* The Social Reaction form included in the Appendix is the one used with the younger children. The form employed with older children was identical except that it included a few additional items.

TABLE 10

TETRACHORIC INTERCORRELATIONS OF RESPONSE ITEMS IN THE CLUSTER, "COOPERATION WITH AUTHORITY"

(Social Reaction Interview,* Older Sample; $N=50$)

Response Number	Response Item	% Frequency of Response	Response Number 3*a*	5*a*	5*b*	7*a*	8*c*	9*a*	1*c*	1*b*	4*a* (School)
Aggression											
3*a*	Remind father if he promised to take him to circus and then didn't	18	—								
5*a*	Play by himself if child visited his house and everybody played with this child and not him	20	.18	—							
Submission											
5*b*	Give toy to other child if mother tells him to	76	.48	−.14	—						
7*a*	Go last if friends say his turn is last	77	.15	.34	.42	—					
8*c*	Not play with child if mother says not to	61	.20	.47	.05	.25	—				
9*a*	Not play with other children if they say he can't	33	.30	−.05	.27	.55	.50	—			
Toilet											
1*c*	Tell adult if child wets his pants	35	.26	−.10	?**	.47	.33	.10	—		
Nudity											
1*b*	Ask mother later if she is taking a bath	55	−.08	.27	−.06	.22	.23	.34	.25	—	
School											
4*a*	Tell teacher if a child scribbles on his paper	74	−.04	.40	.25	.10	.37	.32	.26	.33	—
Assertion											
4*a*	Tell policeman if he is in a strange place and can't find mother	52	.45	.15	.20	.00	.57	.34	.13	.20	.20

* The Social Reaction form included in the Appendix is the one used with the younger children. The form employed with the older children was identical except that it included a few additional items.

** A "?" signifies that the frequencies of the two responses were such as to make the estimation of their coefficient quite uncertain.

the cluster for the younger children. Thus it appears that the *Cooperation with Authority* cluster that we have found for our older sample could be related to the *Socialized Compliance* cluster found for our younger children. The general implications are similar; there is a marked overlap of content, and the differences for the most part may be related to differences in age.

We see that differences in age appear to result in both the disappearance of responses that provide a useful frame of reference for younger children and the appearance of new patterns of responses which are descriptively relevant for the older group. This shift is not surprising and it would be of considerable interest to make sequential studies to illustrate the course of this change. At present, however, we merely see evidence for the change and can relate it to a differential availability and appropriateness of the forms of behavior involved in the Social Reaction Interview changes. The nature of the continuity, if any, and the cause of the change are not clear. Possibly the shift in response pattern is a result of changes in the availability of the various forms of response; perhaps it is a result of differences in specifically taught values.

The differences between the clusters for our younger and older samples are of only incidental interest to us at present. We are primarily interested in the various clusters because they permit us to make rational distinctions among the children in our study. As was indicated in our discussion of the data for the younger children, each of the various bits of behavior that comprise the cluster is considered to be an indication of circumstances under which a given general class or form of behavior may occur. We desire to use the clusters as a basis for distinguishing among our children with respect to the number of different situations in which reactions that belong to a given class characterize the child. Distinguishing among children on the basis of the number of different situations under which a class of reaction or form of behavior occurs is a way of measuring the pervasiveness or the generality of a given form of behavior. Accordingly, the clusters provide the basis for a kind of extensity scale; a high score on the scale simply means that the child shows the form of behavior implied by the cluster in numerous situations, a low score that the child is characterized by the implied form or class of behavior in relatively few situations.

Let us now summarize the differences in clusters for the two age samples. The fact that the younger sample differs from the older sample in ways that we would expect them to differ suggests that the differences may be largely due to differences in age, i.e., may reflect changes in age. Responses to the Social Reaction Interview indicate that *Dependence on Adults* is an unimportant basis for distinguishing between older children. It was also found that *Aggressive* attitudes are mentioned relatively infrequently by the older children. These clusters are not a part of our description of the older children. The *Socialized Compliance* cluster found for the younger children is also apparent for the older children. The data provided by the other children, however, provide evidence for a new cluster that is more descriptive of the older children than the *Socialized Compliance* cluster. This new cluster, called *"Goody-Goody"* has implications similar to those of the *Socialized Compliance* cluster. Accordingly, the *Goody-Goody* cluster will be employed in our analyses for the eight-year-olds and we shall not employ the *Socialized Compliance* cluster. The clusters indicative of *Taking an Adult Role* and indicative of *Weak-Avoidant* reactions are not evident in the data for the older children and, accordingly, are dropped from our analysis of the older children. The *Constructive Approach* cluster is not apparent for the older children, but in the data for the older children there has emerged a new cluster that is similar in its implications.

TABLE 11

PRODUCT-MOMENT INTERCORRELATIONS AMONG CLUSTER SCORES

(Social Reaction Interview, Older Sample; $N=81$)

Cluster	Split-Half Reliability*	Cluster 1	Cluster 2
1. Goody-Goody	.59	—	
2. Responsible Attitude	.52	.16	—
3. Cooperation with Authority	.41	.31	.24

* Based on an odd-even correlation corrected by the Spearman-Brown formula.

This new cluster is called *Cooperation with Authority,* and it will be employed in our analysis for the older sample.

Thus we see that for the older sample of children three cluster scores may be drawn from the Social Reaction Interview: *"Goody-Goody," Responsible Attitude,* and *Cooperation with Authority.* For the cluster scores emerging for the older sample, questions of reliability arise as they did from our data provided by the younger sample. There is no satisfactory answer to these questions, but the criterion of internal consistency may be applied by inspecting the intercorrelations among the bits or items. We have provided split-half reliabilities in Table 11, which provides the intercorrelations among the cluster scores for the older sample.

SCORES DESCRIPTIVE OF THE CHILD BASED ON THE MOTHER'S CHILD-REARING PRACTICES INTERVIEW

A lengthy, semistructured interview was conducted with the mother concerning her child-rearing practices. Although very few specific questions were asked by us during the course of this interview, a standard set of general topics was always introduced for discussion. Our interviewers (Ph.D.'s in psychology, trained in psychotherapeutic interviewing) were prepared with a standard set of questions (see Appendix B) which were not asked, but for which answers were sought and recorded during the course of the interview. In practice, the interview was guided and extended until all of the desired information emerged; whenever specific questions were necessary, they were reserved for the end of the interview. Much of the information had to do with the reactions of the child, including the reactions of the child to the mother, to associates, to himself; it also included his fears and problems (Appendix C). These items were utilized as a basis for evaluation of the child's personal or social development as seen by the adoptive mother. Although the data may not be accepted as a simple, direct description of the child's behavior per se, this evaluative device is none the less relevant. After all, a very important phase of the child's personal and social development is that phase which is seen in the eyes of his mother.

Each of the clusters based on the *Child Rearing Practice Interview* comprises a relatively small number of items and in some cases there are several zero correlations. Here again the naming of the clusters is something of an arbitrary matter and the reader should feel quite free to interpret them differently.

The items that comprise Table 12 all refer to responses which frequently are described in clinical discussions as *Compulsive* responses; among these items the number of conspicuous positive interrelations is quite sufficient to encourage one to place them in a common group and to give some credence to the implied common behavioral feature.

In Table 13 the *Aggressive* cluster is seen to be made up of items that are for the most part not only highly interrelated but of items that are almost unmistakable in their implications. It seems

TABLE 12

TETRACHORIC INTERCORRELATIONS OF RESPONSE ITEMS IN THE CLUSTER, "COMPULSIVE REACTION"

(Child-Rearing Practices Interview with Mother, Younger Sample; $N = 100$)

Response Number	Response Item	% Frequency of Response	Response Number					
			B4	B5	B6	C1	C3	6
Feeding								
B4	Desires specific foods	45	—					
B5	Vomits easily	32	.50	—				
Toilet training								
B6	Is sporadically wet during night	48	.15	.19	—			
Cleanliness								
C1	Initiates washing	63	.25	.44	.11	—		
C3	Initiates change to clean clothes	58	.06	.33	.14	.64	—	
Nervous mannerisms								
6	Masturbates	34	.50	.24	.44	−.03	.02	—
9	Has rituals	15	.34	.14	.37	−.05	.27	.10

TABLE 13

TETRACHORIC INTERCORRELATIONS OF RESPONSE ITEMS IN THE CLUSTER, "AGGRESSIVE"

(Child-Rearing Practices Interview with Mother, Younger Sample; $N=100$)

Response Number	Response Item	% Frequency of Response	Response Number M2	M5	M6	C1	C2	C5
Aggression								
A. M2	Calls mother names when angry with her	63	—					
M5	Has temper tantrums when angry with mother	28	.43	—				
M6	Does spiteful things when angry with mother	34	.56	.47	—			
A. C1	Hits other children when angry with them	74	.27	.31	.34	—		
C2	Calls children names when angry with them	59	.40	.03	.28	.61	—	
C5	Has temper tantrums when angry with children	23	.51	.32	.50	.45	.58	—
C6	Does spiteful things when angry with children	26	.57	.31	.72	.39	.54	.60

unlikely that the naming of this cluster would be the basis for controversy.

In Table 14 the *Phobic* cluster comprises items that all describe fear or other types of anxiety or fearful responses. Although the pattern of intercorrelation is not perfectly consistent, it has sufficient consistency to encourage the belief that children differ with respect to the generality of their fears, and that in many children fearfulness in one situation is associated with fearfulness in another.

In Table 15 the *Dependence* cluster comprises only four items, all of which have positive interrelations although they are low. The writer sees a common behavioral component in these items, however, and since they seemingly describe an aspect of children's behavior

TABLE 14

TETRACHORIC INTERCORRELATIONS OF RESPONSE ITEMS IN THE CLUSTER, "PHOBIC"

(Child-Rearing Practices Interview with Mother, Younger Sample; $N=100$)

Response Number	Response Item	% Frequency of Response	Response Number B2	B5	C1	11	12	14	15
Sleeping									
B2	Awakens easily	21	—						
B5	Has nightmares	48	.25	—					
C1	Is afraid of dark	33	.29	.42	—				
Nervous mannerisms									
11	Fears injury	10	.32	.03	.40	—			
12	Fears water or storms	24	.40	.11	.26	.53	—		
14	Fears animals	23	.50	.25	.30	.43	−.05	—	
15	Fears doctors	19	.24	.08	.26	.22	.05	.04	—
16	Has other fears	21	.38	.35	.60	.32	−.12	.33	.00

Bureau Ednl. Psy. Research
COLLEGE
Dated
No

TABLE 15

TETRACHORIC INTERCORRELATIONS OF RESPONSE ITEMS IN THE CLUSTER, "DEPENDENCE"

(Child-Rearing Practices Interview with Mother, Younger Sample; $N = 100$)

Response Number	Response Item	% Frequency of Response	Response Number		
			B1	B2	B3
Independence					
B1	Often requests help on tasks he can do	62	—		
B2	Often demands attention or affection	77	.27	—	
B3	Goes to parents for comfort in difficulties	68	.13	.32	—
B4	Goes to parents about quarrels with children	49	.30	.10	.40

that is of considerable practical importance to parents as well as to others, we are quite willing to include it in our analyses for whatever it is worth.

Table 16, *Suppression of Aggression,* may be interpreted differently by different readers. It will be regarded by some readers as quite similar in its implications to the *Weakness-Avoidance* cluster score which was described in our discussion of the *Social Reaction Interview.* Actually, however, this *Suppression of Aggression* cluster score does imply some important passive, dependent types of reaction which may often characterize children who do not show ordinary aggressive reactions, or at least tend to avoid aggressive responses.

The intercorrelations among these cluster scores for the *Child Rearing Practices Interview* are shown in Table 17. Note that the split-half reliability estimates for these cluster scores are also included.

The psychologist's Child Rearing Practices Interview with the mothers of the older children was quite similar to the interviews with the mothers of the younger children. The essential nature of any difference was due to the presence of additional items of information in the interview of the mothers of the older children. In the following paragraphs we shall re-examine the clusters of items that

TABLE 16

TETRACHORIC INTERCORRELATIONS OF RESPONSE ITEMS IN THE CLUSTER, "SUPPRESSION OF AGGRESSION

(Child-Rearing Practices Interview with Mother, Younger Sample; $N = 100$)

Response Number	Response Item	% Frequency of Response	Response Number						
			B4	A6	B16	M4	M7	F7	C3
Independence									
B4	Goes to parents about quarrels with children	49	—						
Toilet training									
A6	Did not cry at start of toilet training	21	.46	—					
B16	Has cried after soiling at night	36	.35	−.26	—				
Aggression									
A. M4	Cries when angry with mother	69	.44	.23	−.07	—			
M7	Tries to control himself when angry with mother	23	.06	.37	−.03	.00	—		
A. F7	Tries to control himself when angry with father	31	.34	.50	−.10	−.03	.80	—	
A. C3	Cries when angry with children	59	.50	−.06	.33	.62	.21	.20	—
C4	Sulks or withdraws when angry with children	43	.69	.00	.43	.17	.25	.40	.43

TABLE 17

PRODUCT-MOMENT INTERCORRELATIONS AMONG CLUSTER SCORES

(Child-Rearing Practices Interview with Mother, Younger Sample; $N=114$)

Cluster	Split-half Reliability*	Cluster 1	2	3	4
1. Compulsive	.52	—			
2. Aggressive	.70	.35	—		
3. Phobic	.57	.34	.28	—	
4. Dependence	.71	.04	.17	.06	—
5. Suppression of Aggression	.36	−.11	−.09	−.21	.41

* Based on an odd-even correlation corrected by the Spearman-Brown formula.

were descriptive of the child and based on items from the younger group; we shall consider the manner in which the original clustering observed for the younger children obtains or fails to obtain in the data for the older children. It seems most likely that the descriptive principles (clusters) that were useful in evaluating our younger sample may be different from those most relevant for the older sample.

Let us begin with the seven items that form a cluster of *Compulsive* reactions for the younger sample of children. In general, it appears that these seven items may be slightly less descriptive of the older sample and the intercorrelations indicate that these compulsive items are not consistently related for the older group of children.

In Table 18 we find the intercorrelations for the seven items that comprise an *Aggressive* cluster score for the older sample. The correlations for the older sample shown in Table 18 are positive and indicate that aggressive behavior as seen by the mothers is as consistent for the older children as it was for the younger children. It is noted that some of the aggressive reactions are reported less frequently for the older children than they were for the younger children.

Intercorrelations for *Phobic* items comprised a cluster in the data provided by the mothers of the younger children. The intercorrelations for the older children are less consistent in their implications than those provided for the younger children, however, and the *Phobic* cluster is not considered to be suitably descriptive for the older sample.

The *Dependence* items are less prevalent among the older children than they are among the younger children, nevertheless, they tend to be positively interrelated for the older group in a manner comparable with the younger group. It is not surprising, however, that the kind of dependent behavior that is described by mothers of the five-year-old children should be

TABLE 18

TETRACHORIC INTERCORRELATIONS OF RESPONSE ITEMS IN THE CLUSTER, "AGGRESSIVE"

(Child-Rearing Practices Interviews for Younger and Older Samples)

Parts of Child-Rearing Practices Form	Response Items Based on Cluster for Younger Children	Frequency of Response: % of younger group ($N=114$)	% of older group ($N=81$)	Tetrachoric Correlations Based on Data for 50 Older Children: M2	M5	M6	C1	C2	C5
Aggression									
AM2	Calls mother names when angry with her	63	43	—					
AM5	Has temper tantrums when angry with mother	28	19	.52	—				
AM6	Does spiteful things when angry with mother	34	23	.34	.39	—			
AC1	Hits other children when angry with them	74	73	.15	.01	.34	—		
AC2	Calls children names when angry with them	59	70	.57	.44	.29	.24	—	
AC5	Has temper tantrums when angry with children	23	23	.30	.72	.37	.23	.52	—
AC6	Does spiteful things when angry with children	26	22	.38	.41	.63	.32	.50	.40

TABLE 19

TETRACHORIC INTERCORRELATIONS OF RESPONSE ITEMS IN THE CLUSTER, "ANXIETY"

(Child-Rearing Practices Interview with Mother, Older Sample; $N=50$)

Response Number	Response Item	% Frequency of Response	Response Number							
			B4	B5	B1	B2	A5	B1	6	15
Feeding										
B4	Desires specific foods	29	—							
B5	Vomits easily	29	.36	—						
Independence										
B1	Often requests help on tasks he can do	42	.35	.35	—					
Sleeping										
B2	Awakens easily	23	−.07	.03	.11	—				
Sex										
A5	Participated in sex play with other child (before age of 6 years)	22	.33	.06	.23	.40	—			
Ethics										
B1	Often lies	37	.51	.45	.37	.30	.16	—		
Nervous mannerisms										
6	Masturbates	19	.34	.15	.37	.08	.32	.10	—	
15	Fears doctors	10	.19	.70	.52	.12	.13	.33	.20	—
X (includes 2, 3, 4, 7, 8, 9, 10)	Has rituals or mannerisms	37	.51	.39	.43	.05	.48	.48	.28	.33

somewhat less prevalent in the responses of mothers of the older children.

In the foregoing paragraphs we see that many of the forms of behavior that are descriptive of the younger children are somewhat less conspicuous in the older children. In addition, it is seen that with respect to some behavioral facets responses described by the mothers have a different organization for the older children. Consistencies that are evident for the younger children are rendered doubtful or obscure in the data provided by the older age groups.

The intercorrelations for the older group of children were scrutinized for plausible consistencies, and certain clusters emerged. In some instances the new clusters may have no parallel in our study of the data for the younger children, and in other instances they may be seen as a possible modification or extension of a pattern or response that characterizes the younger children.

In Table 19 we find that there is a consistent interrelationship among several responses which may be considered as indicative of *Anxiety*. The appearance of this cluster in the data for the eight-year-old group and the appearance of the *Phobic* cluster in the data for the younger group provides an interesting contrast and suggests a trend that should receive further scrutiny. Specifically, it suggests that among younger children fears are generalized, so that the child who is fearful in one situation tends to be fearful in other situations. Among older children, however, specifiable fears tend to be less generalized; instead, there is a suggestion that anxiety-type responses (or at least responses that could be interpreted as indicative of anxiety) are generalized.

In Table 20 we find that the data provided by the older sample include 11 items that are consistently interrelated and that appear to have a common behavioral significance. Although these items were somewhat more prevalent among the younger children than among the older children, they were not as consistently organized among the younger children. All of these items could be descriptive of what the layman refers to as a *"Spoiled Child."* The appearance of this cluster indicates important differences among parents. Although the cluster itself can be used to describe children, the behavior of the child for the most part has a particular relevance for the parents. These responses could be seen, in general, as descriptive of aggression in one form or another directed toward the parent, and they imply that some parents are conspicuously unable to reduce this aggressive drive or to divert it in some satisfactory manner.

In Table 21 we find the intercorrelations for 11 items that were not consistently interrelated for the younger children but that are quite consistent in their relationship for the older children. These items refer to the early reactions of the child to his excretory functions, and to the later reaction of the child to certain interpersonal situations. The underlying implication of these items, regardless of the chronological reference, appears to be *Control* of impulses both excretory and interpersonal.

In summary, the psychologist's interview with the mother concerning her child-rearing practices indicated that the responses of children in the older group were organized differently from the responses in the younger group. As a consequence, certain differences in the cluster scores will be employed in our analyses for the older children. The *Compulsive* cluster that characterized the younger children did not characterize the older children. The *Aggressive* responses form a conspicuous cluster for both groups of children, and, accordingly, will be employed in our analyses for the older children. The *Phobic* cluster for the younger children is not apparent in

TABLE 20

TETRACHORIC INTERCORRELATIONS OF RESPONSE ITEMS IN THE CLUSTER, "SPOILED CHILD"

(Child-Rearing Practices Interview with Mother, Older Sample; $N = 50$)

Response Number	Response Item	% Frequency of Response	Response Number									
			B1	B2	B3	M1	M4	M5	M12	F5	C3	A3
Independence												
B1	Often requests help on tasks he can do	42	—									
B2	Often demands attention or affection	66	.50	—								
B3	Goes to parents for comfort in difficulties	51	.23	.36	—							
Aggression												
A-M1	Tries to strike mother when angry at her	11	.44	.10	.33	—						
A-M4	Cries when angry at mother	71	.24	.38	.35	.40	—					
A-M5	Has temper tantrums when angry at mother	19	.28	.38	.02	.57	.59	—				
A-M12	Defends himself by counter-accusations when angry at mother	49	.35	.45	.13	.07	.52	.35	—			
A-F5	Cries when angry at father	66	.23	.35	.48	.27	.87	.37	.32	—		
A-C3	Cries when angry at children	64	.01	.34	.42	.30	.41	.08	.05	.68	—	
Cleanliness												
A3	Played with feces	24	.08	.29	.38	.06	.30	.26	.35	.20	.23	—
Nervous mannerisms												
6	Masturbates	19	.37	.38	.20	.85	.42	.49	−.11	.50	.40	−.07

Bureau of Ednl & Psyl. Research
(S. C. E. R. T.)
9.3.81
J737

TABLE 21

TETRACHORIC INTERCORRELATIONS OF RESPONSE ITEMS IN THE CLUSTER, "CONTROL"

(Child-Rearing Practices Interview with Mother, Older Sample; $N = 50$)

Response Number	Response Item	% Frequency of Response	Response Number									
			A17	B2	B3	B5	B6	B12	B13	B14	C1	C3
Independence												
A17	Assists in entertainment of guests	67	—									
Toilet												
B2	Usually dry all day before age of 24 months	40	.30	—								
B3	Usually dry all night before age of 36 months	55	.07	.62	—							
B5	Ceased to have sporadic accidents during day before 5 years	75	.26	.40	.55	—						
B6	Ceased to have sporadic accidents during night before 5 years	73	.40	.22	.75	.61	—					
Aggression												
B12	Rarely shows anger toward mother	55	.42	.20	.36	.57	.14	—				
B13	Rarely shows anger toward father	73	.32	.39	.29	.59	.24	.81	—			
B14	Rarely shows anger toward other children	45	.54	.33	.33	.58	.25	.71	.55	—		
Cleanliness												
C1	Initiates washing procedures	66	.20	.40	.44	.12	.23	.50	.38	.22	—	
C3	Initiates change to clean clothes	61	.23	.25	.35	.55	.19	.41	.48	.03	.73	—
D5	Uses knife properly for cutting food	47	.04	.08	.55	.22	.21	.45	.12	.25	.28	.15

the behavior described for the older children; there emerges a somewhat related cluster which we shall employ as a basis for an *Anxiety* cluster score. Although *Dependence* items seem less characteristic among the older children than among the younger children, we shall retain the *Dependence* cluster and use it in our analyses for the older group. A *Suppression of Aggression* cluster appears to have no counterpart for the older sample. In addition to the *Anxiety* cluster, other new clusters emerge for the older group and will be employed in our analyses. Among these are the *Spoiled Child* cluster and the *Control* cluster. In analyzing the data for the older children we found the *Aggression* items and the *Anxious* items to be very consistently interrelated (as were the *Aggressive* items and the *Phobic* items for the younger children). We shall combine the *Aggressive* and *Anxious* items to form an *Anxious-Aggressive* cluster and shall conduct some of our analyses with scores based on this particular cluster.

In the foregoing discussion of the *Child-Rearing Practices Interview* data, we have described interrelationships among items of information that were descriptive of the child but were provided by the mother in a semistructured interview situation. Scores based on these clusters cannot be taken as a literal description of the children; instead, they are a description of the children as reported by the mothers. Since they are based on consistent interrelationships, it cannot be claimed that they are capriciously determined and have no general significance. It is possible that to some degree they describe stereotyped attitudes of adoptive mothers toward their children; it seems probable, however, that they also describe stereotyped aspects of the child's reaction as seen by

TABLE 22

PRODUCT-MOMENT INTERCORRELATIONS AMONG CLUSTER SCORES
(Child-Rearing Practices Interview with Mother, Older Sample; $N=81$)

Cluster	Split-half Reliability*	Correlation with Age at Follow-up	Cluster 1	Cluster 2	Cluster 3	Cluster 4
1. Aggressive	.64	−.40	—			
2. Anxiety	.56	−.32	.57	—		
3. Dependence	.60	−.27	.18	.26	—	
4. Spoiled Child	.78	−.35	.46	.42	.68	—
5. Control	.78	.23	−.35	−.36	.11	−.15

* Based on an odd-even correlation corrected by the Spearman-Brown formula.

the adoptive mother. Since the behavior of the child as seen by the mother is no less important to us than the behavior of the child per se, the fact that mothers' attitudes or patterns of thinking are involved to some unknown degree in these clusters does not detract from our interest in them or from their relevance to our purposes. It is important, however, that the reader not overlook the fact that we are describing for the child patterns of behavior as they emerge from the interviews with mothers, and that these patterns of described behavior cannot be divorced from patterns of thinking on the part of the mothers.

In Table 22 the reader will find intercorrelations for various cluster scores that are used in our analyses for the older children. Although these correlations are not especially high (we would not be justified in considering the various clusters separately if they were), the nature and the direction of these correlations is plausible and does not detract from our interpretation of the clusters. Since the older group is somewhat more heterogeneous with respect to age than the younger group and since we consider some of the clusters to have a particular relevance to the age of the child, e.g., our data suggest that the very nature of the cluster may be a function of the child's age, the correlations between the cluster scores and the age of the children at the time of follow-up have been included in Table 22. It will be noted that the most conspicuous correlations are the negative ones relating certain asocial or aggressive behavior on the part of the child with the age of the child. The implication that broadly generalized aggression may be less characteristic of older children than it is of younger children is not only consistent with inferences we draw from our survey of cluster changes, but is also consistent with kinds of changes that are observed to reflect the values of our society.

SCORES DESCRIPTIVE OF THE ENVIRONMENT BASED ON THE MOTHER'S CHILD-REARING PRACTICES INTERVIEW

After our first direct contact with the mother, made by our social worker, the mother and the child later came to our offices. During most of this call, one of the psychologists worked with the child while another psychologist talked with the mother. The interview with the mother usually began with a discussion of child-rearing practices. In general, topics of conversation were introduced in a standard sequence, but the questions for which we sought answers were so clear in our minds that we recognized relevant information regardless of the order in which it was presented. (As the reader may have anticipated, the interviewing procedures and other data-gathering techniques that we have employed were developed through a rather lengthy series of preliminary explorations and pretestings. In the preliminary work we used children from nursery schools and cooperating parents as subjects. Throughout the course of our data-gathering period we had frequent staff meetings wherein we presented and evaluated all the data for a given child. In this way we established and maintained a high degree of uniformity in our techniques and judgments. As the reader may have noted, the nature of our procedure for forming clusters is such that items which were infrequent, poorly defined, or inaccurately observed dropped out, i.e., could not yield a pattern of correlation which would lead to their inclusion in a cluster.)

Much of the information concerning the adoptive family's child-rearing practices was combined to form cluster scores in the same manner as some of the Social Reaction Interview material from the children was combined to form cluster scores. The procedure was to intercorrelate the numerous available bits of information concerning the child-rearing practices of the adoptive home, to scrutinize these intercorrelations in order to select out relatively discrete clusters of items, and then to refine these empirical clusterings on the basis of rational considerations. As a consequence, the items remaining in any cluster are not only positively related with each other and relatively independent of items belonging in some other cluster, but they all imply some common circumstance or common policy or have a common behavioral implication.

The nature of the cluster scores that we derived from the bits of *Child-Rearing Practices Interview* information may be examined by referral to Tables 23 to 26. The selection of items for original intercorrelation was based on frequency of response. Bits of information that occurred very infrequently in our sample or which characterized all the families were excluded at the outset because they obviously could provide no useful basis for distinctions.

In examining the intercorrelations of the items which comprise the various clusters, the reader should remember that each of these correlations is a sampling statistic and, accordingly, is subject to sampling variation. The amount of variation that should be expected among tetrachoric correlation coefficients is greater than that expected for Pearson *r's*. The present tetrachoric coefficients were determined by an employment of computing diagrams. In brief, the reader is advised not to ascribe any particular significance to uniquely high correlations nor to become impressed with the occasional zero or insignificantly small negative correlation coefficients.

Since the labeling of these cluster scores is a discretionary matter and depends not only on the writers' interpretation of the implications of the clusters but also on their choice of words to represent their interpretation or inference, the reader is urged to scrutinize the items comprising the clusters and to draw his own inference concerning their possible behavioral implications. Nevertheless, the writers feel that the variation in home environments, particularly with respect to the role a child may play and the models of behavior provided for him, are expressed to an important degree by the implications that inhere in the clusters that accompany the present discussion. For example, Cluster A, which is called *Eagerness*, might very well be described differently. Nevertheless, the writers see in the items gathered together

TABLE 23

TETRACHORIC INTERCORRELATIONS OF RESPONSE ITEMS IN THE CLUSTER A, "EAGERNESS"

(Child-Rearing Practices Interview with Mother, Younger Sample; $N = 100$)

Response Number	Response Item	% Frequency of Response	Response Number						
			A1	A2	A7 (Independence)	Da	A7 (Toilet)	A8	B10
Feeding									
A1	Infant was awakened for bottle feeding	17	—						
A2	Hungry infant was required to wait for scheduled bottle	36	.72	—					
Independence									
A7	Child does not play outdoors without adult supervision	49	.56	−.07	—				
Da	Child is repeatedly warned to avoid harmful situations	38	.53	.16	.48	—			
Toilet training									
A7	Toilet training was interrupted after it was begun	18	.66	.42	.05	.00	—		
Aggression									
A8	Mother gives in when child is angry with her	11	.21	−.15	.15	.40	.47	—	
B10	Parents reason with child about aggression	65	.34	.48	−.10	.24	.14	−.02	—
Sex training									
A3	Child has frequently observed unclothed adults	35	.31	.43	.17	.12	.35	.02	−.05

in Table 23 a common feature that has to do with a more than usual concern on the part of the mother for the child's development and training. It seems that, within the means available to all mothers, the ones who are characterized by these items have sought to do whatever they could to carry out their responsibility for the child and to take whatever steps are necessary for the safety and physical and spiritual peace of the child, even though it may come at the mother's expense. Perhaps the eighth item in this cluster may be viewed as paradoxical, yet the writers feel that this item may be seen to have a place in this cluster; possibly parents who can live out a wholehearted concern for the child's well-being are not likely to desire to hide or conceal themselves from the child. Perhaps this interpretation is farfetched and uncalled for; nevertheless, the intercorrelations tend to support it.

Cluster B1 (Table 24) is described as comprising practices that reveal a lack of sympathy on the part of the parents. This is called an *Unsympathetic Parent* cluster. The items suggest that some parents seemingly cannot in general understand or feel the child's hopes, pains, or frustrations, or at least that some parents are in general unconcerned with these experiences of a child, or perhaps choose to act as if such things did not exist. At one time we were inclined to call this the Spartan constellation of child-rearing practices; there is no indication that parents who show these positively intercorrelated items yield out of consideration for the child's feelings. This group of responses appears to indicate a rather inflexible child-training attitude on the part of the parents. One wonders how much frustration on the part of the child is involved. Cluster B1 was seen to comprise several parts which in view of some beliefs could have different implications for the development of the child's character. Several of the items included in Cluster B1 have to do with toilet training. Since *Severe Toilet Training* is currently given a rather cherished position

TABLE 24

TETRACHORIC INTERCORRELATIONS OF RESPONSE ITEMS IN THE CLUSTER B, "UNSYMPATHETIC PARENT"

(Child-Rearing Practices Interview with Mother, Younger Sample; $N=100$)

Response Number	Response Item	% Frequency of Response	Response Number									
			D1c	D5	C1c	C2b, c	C3c	A4	C5	B1	D4 (Earlier)	D4 (Now)
Feeding												
D1c	Child is coerced to eat	38	—									
D5	Child wanted bottle at time it was discontinued	24	.11	—								
Independence												
C1c	Child's pain is ignored or reproved	40	−.08	−.05	—							
C2b, c	Child is not soothed or reassured when lonely or afraid	33	.21	.58	.44	—						
C3c	Child is reproved for errors	37	.20	.34	.29	.41	—					
Toilet training												
A4	Child was kept on toilet at least 10 minutes in early toilet training	32	.37	−.13	.05	.40	−.14	—				
C5	Child's messes were disgusting to mother	38	.32	.15	.26	.22	−.08	.37	—			
Aggression												
B1	Child is spanked for aggression toward mother	56	.50	.13	.36	.44	.03	−.05	.38	—		
Cleanliness												
D4	Child learned to use fork at 5 years or earlier	11	.31	.41	.27	.41	.34	.30	.07	.37	—	
D4	Child uses fork now	13	.39	.35	.48	.15	.40	.54	.38	.27	?*	—
Sex training												
B	Child has received severe sex training	41	.30	.34	.04	.25	−.09	.50	.48	.20	.25	.32

* A "?" signifies that the frequencies of the two responses were such as to make the estimation of their coefficient quite uncertain.

TABLE 25

TETRACHORIC INTERCORRELATIONS OF RESPONSE ITEMS IN THE CLUSTER B_2, "SEVERE TOILET TRAINING"

(Child-Rearing Practices Interview with Mother, Younger Sample; $N=100$)

Response Number	Response Item	% Frequency of Response	Response Number				
			A3	A4	B8	B9	C5
Toilet training							
A3	Child was first placed on toilet at age of 1 year or earlier	36	—				
A4	Child was kept on toilet at least 10 minutes in early toilet training	32	.12	—			
B8	Child was ridiculed for soiling	33	.09	.17	—		
B9	Child was hit for soiling	34	.35	.16	.47	—	
C5	Child's messes were disgusting to mother	38	.44	.37	.17	.31	—
Sex training							
B	Child has received severe sex training	41	.23	.50	.32	−.07	.48

in developmental psychology, those items were set aside to provide the basis for a separate cluster. Accordingly, in our analyses we will handle Clusters B1 and B2 separately. Cluster B1 indicates a generally unsympathetic, unfriendly attitude of child rearing and Cluster B2 (Table 25) is specific with respect to rigidity in matters concerning toilet training.

Cluster C (Table 26) is rather an interesting cluster because it very likely is confounded with the mother's report of behavior of the child. Accordingly, in our analyses we shall have to regard with considerable caution the *Punishment for Aggression* cluster in the Child-Rearing Practices and the *Aggression* cluster from the same source. The interrelationship between these two clusters is not clear. One could be an artifact with respect to the other. It is interesting to observe, however, that the various items comprising this *Punishment for Aggression* cluster are rather highly intercorrelated. This suggests that with respect to the mother's attitude or behavior there is some consistency in this area; this certainly agrees well with our subjective impressions based on interviews with the mothers. There are numerous mothers who will not tolerate aggressive expressions by their children in any form.

The intercorrelations for these various cluster scores are shown in Table 27. As we would expect, the scores based on the B1 cluster *(Unsympathetic Parent)* are rather highly correlated with the *Severe Toilet Training* cluster score; they overlap somewhat in content. The rest of the cluster scores concerning the mother's child-rearing practices, however, are relatively independent of each other. The number of items comprising each of the clusters is modest, but for the most part the intercorrelations among items are sufficiently high to encourage us in our expectations that the measures of the Child-Rearing Practices which these cluster scores provide could prove useful to us in our analyses.

It is of interest to examine the data provided by the mothers of the older children and to compare the organization of their responses with the organization of the responses of the mothers of the younger children. The interviews with the mothers of the older children

TABLE 26

TETRACHORIC INTERCORRELATIONS OF RESPONSE ITEMS IN THE CLUSTER C, "PUNISHMENT FOR AGGRESSION"

(Child-Rearing Practices Interview with Mother, Younger Sample; $N = 100$)

Response Number	Response Item	% Frequency of Response	Response Number					
			B1	B2	B3	B4	B5	B6
Aggression								
B1	Child is spanked for aggression toward mother	56	—					
B2	Child is reprimanded for aggression toward mother	70	.24	—				
B3	Child is deprived of privilege for aggression toward mother	63	.17	.29	—			
B4	Child is spanked for aggression toward father	37	.71	.24	.33	—		
B5	Child is reprimanded for aggression toward father	63	.31	.61	.49	.40	—	
B6	Child is deprived of privilege for aggression toward father	48	.26	.46	.81	.47	.61	—
B7	Child is spanked for aggression toward children	20	.53	.22	.33	−.04	.44	.20

included some information that was not available for the younger children. This fact, combined with the fact that in general parents treat older children somewhat differently from younger children, adds interest to our comparison and leads us to the expectation that the differences may contribute to our knowledge of child-rearing practices as they pertain to the age of the child.

The interviews with the mothers of the younger children had provided eight bits of information that were consistently related and were judged by us to indicate an *Eagerness* on the part of the adoptive mother for many phases of child care. There is no indication of the existence of this cluster among the bits of information provided by the mothers of the older children, however. Some of the relevant information concerning feeding and toilet training could not be reliably produced by many of the mothers of the older children and as a consequence did not feature in our analysis. The remaining bits of information which had comprised the *Eagerness* cluster were not systematically interrelated in the data for the older children.

TABLE 27

PRODUCT-MOMENT INTERCORRELATIONS AMONG CLUSTER SCORES

(Child-Rearing Practices Interview, Younger Sample; $N = 114$)

Cluster	Split-half Reliability*	Cluster		
		1	2	3
1. Eagerness	.57	—		
2. Unsympathetic Parent	.44	−.02	—	
3. Severe Toilet Training	.44	−.01	.56	—
4. Punishment for Aggression	.78	−.05	.36	.23

* Based on an odd-even correlation corrected by the Spearman-Brown formula.

We find in the responses of the mothers of the older children, however, a cluster with similar implications. In Table 28 correlations are revealed for a number of items which are considered to be indicative of an overcontrolling attitude on the part of the mother. Since the *Overcontrolling* cluster for the older children has but one item of information in common with the eagerness cluster for the younger children we cannot

TABLE 28

TETRACHORIC INTERCORRELATIONS OF RESPONSE ITEMS IN THE CLUSTER, "OVER CONTROLLING"

(Child-Rearing Practices Interview with Mother, Older Sample; $N = 50$)

Response Number	Response Item	% Frequency of Response	Response Number					
			21a	B5	Da	C3	C5	B9
Independence								
A21a	Mother gives in to child's repeated demands	32	—					
B5	Parent intervenes in child's quarrels with other children	34	.27	—				
Da	Child is repeatedly warned to avoid harmful situations	41	.04	.28	—			
Toilet								
C3	Mother looks at child's stool	15	.22	.60	.30	—		
C5	Child's messes disgusting to mother	27	.40	.38	.30	.48	—	
Aggression								
B9	Child is deprived of privileges for aggression toward children	30	.17	.29	−.03	.49	.32	—
Sleeping								
D4	Child has night light on	41	.25	.00	.08	.20	.37	.05

assume that they are identical or that they necessarily bear a sequential relation to each other. The similarity of their implication for the mother, however, is noteworthy.

In their interviews the mothers of the younger children had provided many bits of information that seemingly pointed toward an *Unsympathetic* reaction to the needs and problems of young children. In our sample of mothers of older children some of the bits were too infrequent to justify analysis.

The mothers of the younger children had also indicated in their responses that there were systematic differences among them with respect to the severity of toilet-training practices. The intercorrelated items of information that provided a *Severe Toilet Training* cluster for the younger children were found to be similarly interrelated in the information from the mothers of the older children (see Table 29). This systematic patterning of the responses of mothers with respect to toilet training leads us to hope that major differences in toilet training may be studied for both the younger and the older children in our study.

In the responses of the mothers of the younger children we found indication that aggressive expression on the part of the children was generally tabooed by some of the mothers. Systematic differences in punishment for aggression were also revealed in the material supplied by the mothers of the older children. As a matter of fact, it seems possible that punishment for aggression may be more prevalent and more distinct for the mothers of the older children than for the mothers of the younger children. As illustrated in Table 30, certain additional bits of information are added to the *Punishment for Aggression* cluster in the data for our older group.

In Table 31 will be found the split-half reliabilities and intercorrelations of these cluster sources for the older sample.

TABLE 29

TETRACHORIC INTERCORRELATIONS OF RESPONSE ITEMS IN THE CLUSTER, "SEVERE TOILET TRAINING"

(Child-Rearing Practices Interview with Mother, Older Sample; $N=50$)

Response Number	Response Item	% Frequency of Response	Response Number				
			A3	A4	B8	B9	C5
Toilet							
A3	Child was first placed on toilet at age of 1 year or earlier	55	—				
A4	Child was kept on toilet at least 10 minutes in early toilet training	58	−.32	—			
B8	Child was ridiculed for soiling	22	.25	.29	—		
B9	Child was hit for soiling	19	.40	.40	.10	—	
C5	Child's messes were disgusting to mother	27	.32	−.05	.45	.30	—
Sex							
B	Child has received severe sex training	35	.26	.46	.30	.14	.48

Other Scores Descriptive of the Home

In addition to the interview concerning child-rearing practices, the mother was interviewed also with respect to the general organization of the home, including its activities, values, recreations, and general pattern of living. The content of this interview was recorded on a form called the Conditions of Child Rearing; a sample of this is included in Appendix D. Another source of information descriptive of the home was provided by the use of a set of rating scales called the Multiple Rating Scale. The ratings were made by the psychologist as a result of the interview with the mother. A copy of this is included in Appendix E. The Multiple Rating Scale is of unusual interest because it is of such a nature as to be employed by both the social worker who visited the home and the psychologist who interviewed the mother. Accordingly, we have two sets of these ratings for each family. (As we shall see, these two sources of information covering a common phase of the home will be useful in our subsequent concern with a possible confounding between the nature of our findings and the source of our data.) In addition to the Multiple Rating Scale, the social worker used a Home Rating Scale in recording her observations at the time of her visit to the home. Appendix F presents a copy of this scale.

These three forms—Conditions of Child Rearing, Multiple Rating Scale, and Home Rating Scale—provided bits of information descriptive of many important facets of the home other than child rearing per se, and accordingly were used as a basis for forming cluster scores descriptive of the home and its general atmosphere. (At some points we refer generically to these clusters as "Conditions of Child Rearing" cluster scores, although they obviously are not limited to the items included in the Conditions of Child Rearing form.) The procedure for forming these cluster scores is the one described for the other cluster scores: specifically, items for which responses occurred with inappropriate frequency were eliminated; the remaining items were intercorrelated, and those that showed fewer significant, positive correlations than might be expected on the basis of chance were eliminated. On the basis of inspection, relatively discrete clusters of interrelated items were formed; these clusters were further refined by excluding all items that seemed

TABLE 30

TETRACHORIC INTERCORRELATIONS OF RESPONSE ITEMS IN THE CLUSTER, PUNISHMENT FOR AGGRESSION

(Child-Rearing Practices Interviews for Younger and Older Samples)

Parts of Child-Rearing Practices Form	Response Items Based on Cluster for Younger Children	Frequency of Response % of younger group (N=100)	Frequency of Response % of older group (N=50)	Tetrachoric Correlations Based on Data for 50 Older Children B1	B2	B3	B4	B5	B6	B7	B8	C1a
Aggression												
B1	Child is spanked for aggression toward mother	56	42	—								
B2	Child is reprimanded for aggression toward mother	70	84	.30	—							
B3	Child is deprived of privilege for aggression toward mother	63	74	.00	.33	—						
B4	Child is spanked for aggression toward father	37	25	.75	.30	.62	—					
B5	Child is reprimanded for aggression toward father	63	77	−.02	.80	.60	.30	—				
B6	Child is deprived of privilege for aggression toward father	48	69	.15	.68	.92	.45	.70	—			
B7	Child is spanked for aggression toward children	20	13	.40	−.16	−.21	−.04	.16	−.12	—		
	Additional Response Items Based on Cluster for Older Children											
Toilet												
B8	Child was ridiculed for soiling	33	22	.14	.42	.30	.25	.54	.28	.00	—	
Aggression												
C1a	Parents spank child less now than in earlier years	—	68	.32	.31	.45	.28	.55	.63	.15	.20	—
C1c	Parents deprive child of privilege more now than in earlier years	—	59	.02	.26	.66	.35	.59	.77	.04	.18	.40

TABLE 31

PRODUCT-MOMENT INTERCORRELATIONS AMONG CLUSTER SCORES

(Child-Rearing Practices Interview, Older Sample; $N=81$)

Cluster	Split-half Reliability*	Cluster 1	2	3
1. Overcontrolling	.32	—		
2. Severe Toilet Training	.45	.35	—	
3. Punishment for Aggression	.82	−.04	.18	—
4. Unsympathetic Parent	.33	.18	.67	.04

* Based on an odd-even correlation corrected by the Spearman-Brown formula.

to be behaviorally inconsistent with the over-all implications of the group.

The items that have been gathered together to form Cluster D, *Ambition,* are seen to bear consistent positive relations with each other and seem to the writer to involve a respect for achievement (Table 32). It appears that evidences of interest in achievement of various types all tend to go together. Most of the types of achievement that we note here, however, have to do with a relatively high status in our social and economic structure.

Cluster E (Table 33) is called *Harmonious Family Relationships.* Its component items have positive intercorrelations and for the most part show that the members of a family may be congenial and enjoy each other's company in several respects. Some of the items appear twice in this cluster score depending on whether or not they are based on the social worker's observations or the psychologist's observations.[4]

---

[4] For most items, the psychologist's ratings and the social worker's ratings are in good agreement, despite the fact that they were based on observations at both different times and places. Where there is evidence of disagreement, the social worker's ratings and the psychologist's ratings are handled as separate items.

TABLE 32

TETRACHORIC INTERCORRELATIONS OF RESPONSE ITEMS IN THE CLUSTER, "AMBITION"

(Conditions of Child-Rearing Interview with Mother and the Multiple Rating Scale D, Younger Sample; $N=100$)

	Response Item	% Frequency of Response	Response Number									
			8a	13a	19a	21a	22a	23, 4	23, 5	23, 11	23, 15	23, 16
Parts of Conditions of Child-Rearing Form												
8. Community service	a. At least one parent is very active in community service	25	—									
13. Need achievement	a. One or both parents are considered outstanding in the community	26	.87	—								
19. Culture (music)	a. Parents frequently listen to opera or classical music	45	.64	.60	—							
21. Culture (literature)	a. Parents read classics, outstanding novels, literary magazines, etc.	16	.54	.50	.80	—						
22. Culture	a. Parents frequently play musical instruments, paint or sculpture	17	.31	.72	.41	.54	—					
23. Recreation	4. One or both parents frequently travel on pleasure trips	20	.48	.29	.18	.36	.21	—				
Recreation	5. One or both parents frequently go to the legitimate theater	11	.46	.44	.40	.39	.38	.44	—			
Recreation	11. One or both parents collect non-art objects	8	.36	.50	.40	.18	.35	.29	.30	—		
Recreation	15. One or both parents frequently go to lectures	7	.58	.42	.15	.60	.42	.35	.55	—	—	
Recreation	16. One or both parents frequently read	45	.44	.40	.42	.60	.24	.35	.27	−.10	.35	—
Part of Multiple Rating Scale Form												
5. Flexibility of parents' expectations	a. Parents have fixed ideas about child's education and vocation	8	.53	.34	.60	.39	.52	−.10	.30	.42	.66	.40

TABLE 33

TETRACHORIC INTERCORRELATIONS OF RESPONSE ITEMS IN THE CLUSTER E, "HARMONIOUS FAMILY RELATIONSHIPS"

(Conditions of Child-Rearing Interview with Mother and the Multiple Rating Scale, Younger Scale; $N=100$)

	Response Item	% Frequency of Response	Response Number								
			15a	24c	25c	26d	8a	9d	6a*	6a†	7c*
Parts of Conditions of Child-Rearing Form											
15. Story telling	a. Parents and child usually read together	83	—								
24. Mother's original educational and vocational aspirations	c. Mother's original ambitions are in line with present life situation	59	—	—							
25. Father's original educational and vocational aspirations	c. Father's original ambitions are in line with present life situation	57	—	—	—						
26. Family Compatibility	d. Parents are compatible and harmonious	39	.40	—	—	—					
Parts of Home Rating Scale Form											
8. Sociability (visitors)	a. Family has guests frequently	49	.24	—	—	.32	—				
9. Family group recreation	d. Family enjoys group activities	62	.25	—	—	.53	.55	—			
Parts of Multiple Rating Scale Form											
6. Mother's enjoyment of child	a. Mother shows pleasure in child's behavior (Psychologist)	38	.26	—	—	.28	.03	.37	—		
6. Mother's enjoyment of child	a. Mother shows pleasure in child's behavior (Social Worker)	30	.55	—	—	.10	.31	.50	.41	—	
7. Mother's reaction to child	c. Mother is objective but favorable toward child (Psychologist)	48	.22	—	—	.27	.03	.34	.66	.18	—
7. Mother's reaction to child	c. Mother is objective but favorable toward child (Social	[illegible]	[illegible]	—	—	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	.51

Cluster F (Table 34) had been tentatively called a slovenliness cluster. Perhaps this is a harsh designation for this cluster and it probably implies more disapproval than the investigators actually feel. Accordingly, it is now designated as the *No Ambition* cluster. It is certain that these items neither imply the existence of success in any socially approved undertaking nor indicate any striving for such success. These are the kinds of things that customarily are not characteristic of the families and the homes that are respected in the community because of achievement or because of community responsibility.

Cluster G (Table 35) we have called *Rejection of Child.* Perhaps this designation may receive the least sympathy from the reader because "rejection of child" is a concept that has deteriorated with familiarity and to many people has become almost meaningless. Nevertheless, we find here a group of items that bear a rather impressive degree of positive interrelationship and that fail to indicate any friendliness toward the child or any enjoyment from the child's existence. It seems that all of these items tend to distinguish between parents in much the same way, and that the basis of this distinction has to do with the friendliness of the parent toward the child. One could say that the lack of friendliness is due to the mother's dislike of the child, or perhaps it indicates an inability for liking any child. At any rate, it would seem that the distinction which a score based on this cluster would provide could be related to our concept of rejection.

We have in Cluster H (Table 36) a group of items that are positively interrelated and that suggest to us that the parents may not enjoy each other's company, or at least that the parents may fail to experience a great deal of satisfaction

TABLE 34

TETRACHORIC INTERCORRELATIONS OF RESPONSE ITEMS IN THE CLUSTER "NO AMBITION"

(Conditions of Child-Rearing Interview with Mother and Multiple Rating Scale, Younger Sample; $N = 100$)

Response Item		% Frequency of Response	Response Number								
			8c	11c	13c	18c	19c	21c	1c	2c	3c
Parts of Conditions of Child-Rearing Form											
8. Community service	*c.* Both parents are usually inactive in community affairs	55	—								
11. Routine	*c.* Home schedule is poorly organized	10	.43	—							
13. Need achievement	*c.* Both parents expend only as much effort as is required	17	.60	.26	—						
18. Radio (politics)	*c.* Parents rarely or never listen to forums, news analysts, etc.	38	.27	.57	.61	—					
19. Culture (music)	*c.* Parents listen almost exclusively to popular music	38	.46	.45	.35	.50	—				
21. Culture (literature)	*c.* Parents read almost exclusively mysteries, comics, pulp magazines, etc.	20	.37	.48	.52	.40	.54	—			
Parts of Home Rating Scale Form											
1. Physical Plant	*c.* Home does not meet physical need of child	9	.54	.30	.30	.55	.55	.40	—		
2. Outdoor Play Space	*c.* Home has little or no play space	10	.55	.27	.42	.30	.60	.35	.87	—	
3. Maintenance of Home	*c.* Exterior and grounds of home are not well maintained	25	.74	.39	.50	.29	.56	.40	—	—	—
5. Neatness	*c.* Home is characterized by disarray	7	.44	.60	.41	.24	.25	.35	.71	.40	.47

or pleasure in each other's company. Perhaps the emphasis on pleasure in each other's company is unwarranted here and it may be a difference with respect to values other than purely social ones. At any rate, we have chosen to consider this a basis for a measure of *Parental Incompatibility* and we do this with full awareness that the term may be misleading to some of our readers. In every case, however, we encourage our readers to examine the items comprising each cluster and form their own inference concerning its possible meaning. Just as the meaning of a new intelligence test cannot be known until first the examiner familiarizes himself with the contents, the meaning of cluster scores cannot be known till the reader examines the bits of information that go together to form the cluster.

Cluster I (Table 37) is made up of five items which bear rather consistent positive interrelations with each other. Although the word *Overprotective* would seem to be justifiable in view of the nature of the items, perhaps some will feel that the cluster is a little pretentious inasmuch as it really involves only five bits of information that are based on but three different items. It should be noted, however, that the psychologist and the social worker have seen the mother under quite different conditions and, accordingly, their rating of the mother is not based on the same sample of behavior, i.e., the behavior was observed not only by two different people but it was observed in two different contexts. The fact that the social worker's and the psychologist's observations are not identical is indicated by the particular fact that they do not bear an exceptionally high correlation with each other. It is interesting (see Table 38) that the split-half estimates for reliability are relatively high for the

TABLE 35

TETRACHORIC INTERCORRELATIONS OF RESPONSE ITEMS IN THE CLUSTER G, "REJECTION OF CHILD"

(Conditions of Child-Rearing Interview with Mother and the Multiple Rating Scale, Younger Sample; $N = 100$)

Response Item		% Frequency of Response	Response Number							
			9a	14a	16c	1b, c	3c*	3c†	6c, d*	6c, d†
Parts of Conditions of Child-Rearing Form										
9. Emotionality of mother	*a.* When angry, mother usually expresses her feelings without restraint	30	—							
14. Expressed disapproval of parents	*a.* Parents are very critical of child	19	.65	—						
16. Conversation between mother-child	*c.* Mother tends to discourage child from talking	17	.30	−.03	—					
Parts of Multiple Rating Scale Form										
1. Acceptance of child	*b* and *c.* Parents do not completely accept child as their own	32	.47	.37	.45	—				
3. Affectionateness of mother	*c.* Mother rarely fondles or kisses child (Psychologist)	19	.32	.40	.23	.29	—			
3. Affectionateness of mother	*c.* Mother rarely fondles or kisses child (Social Worker)	16	.35	.27	.20	.31	.15	—		
6. Mother's enjoyment of child	*c* and *d.* Mother shows little or no real enjoyment in child's behavior (Psychologist)		.48	.21	.27	.76	.22	.42	—	
6. Mother's enjoyment of child	*c* and *d.* Mother shows little or no real enjoyment in child's behavior (Social Worker)	22	.40	.24	.29	.85	.37	.45	.73	—
7. Mother's reaction to child	*a.* Mother is critical of child, describing his behavior in uncomplimentary ways	13	.67	.41	.47	.75	.25	.13	.75	.70

* Psychologist.
† Social Worker.

TABLE 36

TETRACHORIC INTERCORRELATIONS OF RESPONSE ITEMS IN THE CLUSTER H, "PARENTAL INCOMPATIBILITY"

(Conditions of Child-Rearing Interview with Mother and the Multiple Rating Scale, Younger Sample; $N=100$)

Response Item		% Frequency of Response	Response Number			
			5c	24a	25a	26a, b
Parts of Conditions of Child-Rearing Form						
5. Parental contacts	*c.* Parents rarely spend leisure time together	15	—			
14. Mother's original educational and vocational aspirations	*a.* Mother's aspirations are markedly superior to present situation	7	—	—		
25. Father's original educational and vocational aspirations	*a.* Father's aspirations are markedly superior to present situation	11	—	—	—	
26. Family compatibility	*a* and *b.* Parents are very incompatible—may have considered divorce	11	.67	—	—	—
Parts of Home Rating Scale Form						
9. Family group recreation	*c.* Family rarely enjoys group activities	9	.65	—	—	.47

various cluster scores in this group, despite the fact that many of the cluster scores are based on items from different sources. It will be noted that in some instances a Multiple Rating item from the psychologist's interview and the same Multiple Rating item from the social worker's visit are used as components of a given cluster score. Since the items obviously have a similar meaning and are positively correlated, it is reasonable that they should fall into the same cluster score. The items are not identical, however; they refer to behavior observed by

TABLE 37

TETRACHORIC INTERCORRELATIONS OF RESPONSE ITEMS IN THE CLUSTER I, "OVERPROTECTIVENESS"

(Conditions of Child-Rearing Interview with Mother and the Multiple Rating Scale, Younger Sample; $N=100$)

Response Item		% Frequency of Response	Response Number			
			2a*	2a†	4d	7d*
Part of Multiple Rating Scale Form						
2. Solicitousness towards child	*a.* Mother is oversolicitous with child (Psychologist)	26	—			
2. Solicitousness towards child	*a.* Mother is oversolicitous with child (Social worker)	12	.27	—		
4. Accelerational attempt	*d.* Parent tries to maintain child's immaturity	13	.65	.28	—	
7. Mother's reaction to child	*d.* Mother has marked protective bias toward child's behavior (Psychologist)	14	.30	.07	.38	—
7. Mother's reaction to child	*d.* Mother has marked protective bias toward child's behavior (Social worker)	12	.27	.47	.44	.51

* Psychologist.
† Social Worker.

TABLE 38

PRODUCT-MOMENT INTERCORRELATIONS AMONG CLUSTER SCORES

(Child-Rearing Practices Interview, Younger Sample; $N=114$)

Cluster	Split-half Reliability*	Cluster 1	2	3	4	5
1. Ambition	.78	—				
2. Harmonious Family Relationships	.58	.03	—			
3. No ambition	.72	−.59	−.29	—		
4. Rejection	.76	.14	−.54	.10	—	
5. Parental Incompatibility	.73	.06	−.46	.32	.41	—
6. Overprotectiveness	.77	−.10	−.32	.19	−.01	.20

* Based on an odd-even correlation corrected by the Spearman-Brown formula.

two different persons on two different occasions and in two greatly different situations.

The Rejection cluster score comprised several parts of the Multiple Rating Scale form. Since the Multiple Rating Scale data were provided independently both by the social worker and by the psychologist, we decided to make a special check on the reliability of a "Limited" Rejection cluster score. We did this by forming for each child a Rejection cluster score based on the social worker's observations and a Rejection cluster score based on the psychologist's interview. Each of these partial scores comprised the ratings on Acceptance of the Child, Affectionateness of the Mother, and the Mother's Enjoyment of the Child. Obviously, a cluster score based on these few items would be greatly restricted in range; nevertheless, the respective restricted cluster scores correlated with each other to the order of .60. Since the type of information involved is somewhat judgmental in character and therefore not likely to yield the highest evidence of reliability, and since the basic observations were made under two greatly different conditions, we are inclined to consider this correlation of .60 as an indication that our cluster scores emerging from the various bits of information probably have a useful degree of reliability.

The intercorrelations among the cluster scores are shown in Table 38. Some of these intercorrelations are of considerable interest. As we would anticipate, *Ambition* and *No Ambition* are inversely related. It is also of interest to observe that *Harmonious Family Relationships* is inversely related with our *Rejection of Child* cluster score, and that the *Rejection* cluster score is positively related with the measure of *Parental Incompatibility.* (This confirms the work of others [3,16].) To complete the discussion of these parts, *Parental Incompatibility* is seen to be inversely related with *Harmonious Family Relationships.* Thus we see that the patterns of intercorrelations among the cluster scores coming from observations made by different raters in different contexts are consistent with each other in their implications.

When we consider the respective data for the older sample, the cluster scores that emerge from the material summarized on the *Conditions of Child Rearing* form, *Multiple Rating Scale,* and *Home Rating Scale* (see Appendixes) are quite similar to those emerging for the younger sample. That is to say, the respective clusters are either highly similar in content or quite analogous in their implications. There is one conspicuous exception to the over-all similarity. The *Overprotectiveness* cluster that emerged in the data for the younger children was not discernible in our analyses for the older sample.

In our study of the children and of the children's environment we sought to employ observations and procedures that would afford the greatest possibility of revealing relationships between circumstances of the child's rearing and criteria of the child's follow-up development. We have been guided primarily by what we thought were both probable and important kinds of relationships. We have been guided also by what we thought were the most successful efforts of others, and in every case our efforts can be related to the efforts of others—observational, analytical, or theoretical. Perhaps we may sum up best by saying that we have sought here to use a sampling procedure. We have sought representative information for the kinds of experiences and circumstances which we (on the basis of the consensus of the beliefs of others, and our own experiences as well) suspected would be most important in determining or anticipating our criteria. In such an undertaking as this, failure for a certain relationship to appear cannot be offered as evidence that such a relationship does not exist.

Lack of apparent relationship can be due to many things: the use of unreliable variables, faulty interpretation of a theoretical position, untoward but unrecognized characteristics of the sample, etc. If, however, in more aspects of our analyses than could be predicted from chance alone we find important and plausible relationships, we are inclined to ascribe some significance to them and attempt to relate them to the over-all purposes of our study. Lest we be misunderstood, however, it should be emphasized that we have not sought to sample or measure all of the kinds of environment which could conceivably be relevant. We have been limited by our obligation to the agencies, by the delicate nature of the status of adoptive children and adoptive families, by the time and material limitations imposed on any practicable piece of research, and by any biases of orientation that are based on our scrutiny of the published literature. Within these limitations we have not sought to measure intensively but in general have sought to measure extensively.

SUMMARY

Let us review and summarize the nature of our methodological problem.

1. We desired useful measures of several aspects of the social and emotional development of young children. It is possible to think of at least four major aspects of the child's emotional and social development. One aspect would comprise an objective description of the child's behavior. Another aspect would comprise a description of the child as he is regarded by his associates outside of the home. Another would be the mother's opinion of the child. And a fourth would be the child's opinion of himself. Unfortunately, the attitude of our society toward adoptive families, and adoptive children particularly, is such that it would be very difficult if not impossible to visit schools and playgrounds and observe adoptive children and learn about the opinions of others concerning them. As a matter of fact, getting such data for a widely scattered group of children would be very difficult under any circumstances. Accordingly, we had to restrict the content of our inquiry to the child's opinion of himself and to the mother's opinion of the child. It is ob-

vious, therefore, that we do not have anything that could be called a total study of the child's adjustment. Nevertheless, the child's opinion of his own behavior and what the mother thinks of the child's behavior are very practical factors in the child's personal adjustment.

2. We desired useful measures of several aspects of the physical and psychological environment of our children. We had but two opportunities to get such information. One opportunity came at the time of our social worker's visit to the home. This was her first and only visit and had as its ostensible purpose the arrangement of the adoptive family's visit to the Institute of Human Relations. Our second opportunity to learn about the home came at the time of the mother's visit to our offices; while the child was being examined, a semistructured interview was conducted with the mother concerning the home in general and its child-rearing practices in particular. It was in this informal clinical context that we were obliged to get systematic, quantifiable information concerning the home.

3. The financing and planning of our investigation was of such a nature that no provision could be made for a preliminary period of study wherein measuring devices could be constructed. In effect, our position was a paradoxical one where we were obliged to construct measuring devices during the course of our inquiry. The only alternative was one of greatly restricting the scope of the study so that practically no information concerning the personal and social development of the child and practically no information concerning the psychological environment of the home would be available. Rather than so impoverish our investigation, we applied the procedures described in the present report.

4. Although we were confident that we could secure informational content relevant to the purpose of our investigation, we had the special problem of deciding how such content should be used. We were fearful that any a priori combination of information could be seriously misleading, and we were most desirous of using scores that had a meaning which could be empirically defended. Specifically, we were fearful that if the information were combined on some a priori basis to form scores, a given score might not have the meaning that had been ascribed to it. It was decided, therefore, that the most defensible score to use would be based on positively interrelated bits of information. Presumably such scores could have the same general meaning from individual to individual although the items comprising the identical scores of any two given individuals' scores might not be identical.

5. When one is using untried variables, one must have some assurance that the variables have a minimal useful reliability so that their use in exploratory investigation would not be doomed to failure from the start because of a total lack of reliability. It was our opinion that using positively interrelated bits of information as a basis for scoring would guarantee a certain minimal reliability. Specifically, consistently interrelated bits of information must themselves possess some essential reliability; otherwise they could not be so interrelated.

In the Appendixes one may find sets of forms that were used in order to provide standard records of the kinds of information that would be most relevant to our scoring purposes. In our use of these forms we excluded all items or bits that occurred with a frequency inappropriate for meaningful tetrachoric correlations, and in our study of the intercorrelations among the items we discarded all items that did not show more significant positive correlations than would have been expected on the basis of chance. In order to make our cluster scores as definitive and unambiguous as possible, we used only those items that showed positive interrelationships.

Although our principal analyses were based on the sample of younger children, all the analyses were replicated by the use of a sample of older children. It is interesting to note that the clusters found for the sample of younger children are not always similar to those for the sample of older children. In addition, clusters emerge which are appropriate for the description of the older children but are not appropriate for the description of the younger children. This change in the content of descriptive cluster scores is most apparent in the summary provided by Table 39.

Apparently the Social Reaction Interview may be appropriate enough for learning about some of the attitudes of preschool children. It is obvious that this device yields very little for the older children. In view of the tendency, emphasized in the child's school experience, to present one's attitudes in the best light, this result is not surprising. Table 39 also indicates certain differences between the behavior of older and younger children as reported by their mothers. This is seen to be a reasonable trend and is in keeping with commonly observed differences between children of these ages. Commensurate with the differences in children, there are also differences in the child-rearing practices of the parents. Some of the patterns of child-rearing practices that are apparent in the families' reaction to preschool children are

TABLE 39

CLUSTERS FOR YOUNGER AND OLDER SAMPLES, WITH SPLIT-HALF RELIABILITIES

Younger Sample		*Older Sample*	
Social Reaction Interview			
Dependence on adults	.88	Goody-goody	.59
Aggression	.58	Responsible attitude	.52
Socialized compliance	.70	Cooperation with authority	.41
Taking on adult role	.69		
Weakness-avoidance	.57		
Constructive approach	.58		
Child-Rearing Practices Interview (Child)			
Compulsive	.52	Aggressive	.64
Aggressive	.70	Anxiety	.56
Phobic	.57	Spoiled child	.78
Dependence	.71	Control	.78
Suppression of aggression	.36	Dependence	.60
Child-Rearing Practices Interview (Home)			
Eagerness	.57	Overcontrolling	.32
Unsympathetic parent	.44	Severe toilet training	.45
Severe toilet training	.44	Punishment for aggression	.82
Punishment for aggression	.78	Unsympathetic parent	.33
Conditions of Child-Rearing Interview (and Others)			
Ambition	.78	Ambition	.77
Harmonious family relationships	.58	Harmonious family relationships	.60
No ambition	.72	No ambition	.74
Rejection	.76	Rejection	.68
Incompatibility	.73	Incompatibility	.52
Overprotectiveness	.77		

obscured or absent in the families' reaction to the older children.

It will be noted that the material descriptive of the home per se, i.e., the facilities of the home and its values and activities, has an appreciable uniformity of structure for the two age groups. This is reasonable enough inasmuch as these scores (Conditions of Child Rearing) are not oriented about the child primarily, but are oriented about the home in general and the parents in particular.

In Table 39 it will be seen that many of our cluster scores have an odd-even reliability that is encouragingly high. It is quite possible that this approach, with some refinements, could result in interviewing procedures which yield quantifiable material comparable with the material gathered by testing devices.

It should be noted that the mother's information descriptive of the child and the mother's information descriptive of the home have a common and immediate source, and one could fear that some pattern of interrelationships between such variables could emerge primarily as a result of the common source. Fortunately, information indicative of rejective features of the home was gathered by both the psychologist and the social worker at different times and in different settings. We find an appreciable consistency between these two sources of information. Specifically various aspects of the child's behavior revealed by cluster scores emerging from the interview with the mother showed the same kind of relationships with the rejective information whether it comes from the social worker's visit or whether it comes from the psychologist's interview. It is obvious, therefore, that characteristics of the child as described by the mother and characteristics of the home do not have to have the mother as a common source of information in order for them to be interrelated.

From the beginning, we were aware that some of our information descriptive of the children and descriptive of the home could show a spurious evidence of relationship because in some instances they were recorded by the same interviewer. In order to minimize such a possible source of bias, the interviewers never knew the general nature of the scores that were to emerge from our analysis of the data, nor could they have known the manner in which the various items would contribute to the various scores.

In conclusion, our report indicates that one need not abandon hope of a quantitative approach to the study of human behavior merely because no standard measuring instruments are appropriate to the content of the investigation contemplated. It would appear that if one has means of securing the relevant informational content in standard form, one may use this content as a basis for a quantitative procedure, and accordingly, may enjoy the benefits of meaningful variables relevant to the desired content and possessing sufficient reliability to permit one to discover the nature of the most important relationships among the variables. After the nature of the most important relationships is discovered, one may or may not desire to perfect the measuring devices in such a way as to determine the exact degree of relationship.

Appendix A

SOCIAL REACTION INTERVIEW

DIRECTIONS: Ask the child the question and indicate which of the listed responses his answer most resembles. If his response is not listed, write it under "remarks." For the term "child" substitute "boy" or "girl" according to the sex of the interviewee.

INSTRUCTIONS: Let's play a game of pretending. I'll tell you what's happened and you tell me all the things you would do.

Dependency

1. Let's pretend you just got up in the morning and want to get dressed, what do you do?
   a. Dress self.
   b. Tell adult.
   c. Ask what to wear.
   d. Wait.
   e. Get washed.
   f. Remarks—

2. Let's pretend you are playing with your coat on and get too warm, what would you do?
   a. Take off your coat.
   b. Tell an adult that you are warm.
   c. Ask permission to take off the coat.
   d. Keep on playing.
   e. Remarks—

3. Let's pretend your mother asked you to take a bath by yourself, what would you do?
   a. Go ahead and take a bath.
   b. Ask her to help you.
   c. Tell her you don't need one.
   d. Wait for parent's help.
   e. Say that you can't.
   f. Inquire or acquire bath toys.
   g. Remarks—

4. When you get your meat for supper, what do you do about cutting it?
   a. Cut it.
   b. Ask to have it cut.
   c. Ask for something else.
   d. Refuse it.
   e. Remarks—

5. Suppose you get hungry between meals, what do you do?
   a. Get something.
   b. Tell an adult.
   c. Nothing.
   d. Go to friends.
   e. Eat.
   f. Remarks—

6. Suppose you were playing and your toy broke, what would you do?
   a. Try to, or fix it.
   b. Ask for help to fix it.
   c. Get another toy.
   d. Tell an adult.
   e. Stop playing.
   f. Remarks—

7. Suppose you're playing with a friend and he has to leave, what do you do?
   a. Keep on playing.
   b. Stop playing.
   c. Ask him to stay.
   d. Look for another friend or activity.
   e. Ask him whether he will return.
   f. Say goodbye.
   g. Remarks—

8. Suppose you fell down and hurt yourself, what would you do?
   a. Forget about it.
   b. Cry.
   c. Tell an adult.
   d. Wash it off, bandage.
   e. Remarks—

9. When you have to go to the bathroom what do you do?
   a. Go.
   b. Tell an adult.
   c. Wait.
   d. Soils self.
   e. Preparatory behavior.
   f. Remarks—

10. When you're finished on the toilet, what do you do?
   a. Completes toilet activities alone.
   b. Request assistance
   c. I tell or show mother.
   d. Remarks—

11. Suppose your mother asked you to go to the store for her, what would you do?
a. Go.
b. Ask for company.
c. Say "no."
d. Ask to be permitted to get something for myself.
e. Statement of inability to go to store alone.
f. Do something else.
g. Remarks—

12. Suppose your mother has to go away for awhile, what would you do?
a. Stay with another adult.
b. I'd want to go also.
c. Ask her for a present when she returns.
d. Say I don't want her to go.
e. Say goodbye.
f. I'd stay by myself.
g. I'd wait for her.
h. Remarks—

Aggression

1. Suppose you wanted your friend's toy, what would you do?
a. Take it forcibly.
b. Tell an adult.
c. Ask him.
d. Give him mine.
e. Take it surreptitiously.
f. Remarks—

2. If a child won't let you play with his toys, what would you do?
a. Hit him.
b. Play with them anyway.
c. Don't let him play with mine.
d. Say he's selfish.
e. Tell an adult.
f. Offer to let him play with my toys.
g. Play with own toys.
h. Don't play with him.
i. Remarks—

3. Suppose your father promised to take you to the circus and then didn't, what would you do?
a. Remind him of his promise.
b. Cry.
c. Play with toys or friends.
d. Tell mother.
e. Go anyway.
f. I wouldn't go.
g. Remarks—

4. Suppose you wanted ice cream and your mother said you couldn't have it, what would you do?
a. Tell her I want it.
b. Ask for candy, bread and butter.
c. Ask father.
d. Do or say something angry.
e. I'd get it anyway.
f. I wouldn't have it.
g. Remarks—

5. Suppose a child visited at your house. Everybody played with the child and nobody played with you, what would you do?
a. Play by myself.
b. Cry.
c. Go away.
d. Hit the child.
e. I'd go to my mother.
f. Tell child to leave.
g. Make bid for attention.
h. Remarks—

6. Suppose a child hits you on purpose, what would you do?
a. Hit him back.
b. Tell an adult.
c. Cry.
d. Tell him not to hit you.
e. Go away from him.
f. Give him what he wants, try to please him.
g. Remarks—

7. Suppose a child takes a toy of yours that you don't want him to have, what would you do?
a. Hit him.
b. Tell an adult.
c. Take it back.
d. Tell him to give it back.
e. Take something of his.
f. Let him have it.
g. Remarks—

8. If another child calls you a bad name, what would you do?
a. Hit him.
b. Tell an adult.
c. Say something bad to him.
d. Pay no attention to it.
e. I'd say I'm not bad.
f. I wouldn't play with him.
g. Remarks—

Submission

1. When your mother tells you to eat something you don't like, what do you do?
    - *a.* Don't eat it.
    - *b.* Eat it.
    - *c.* Eat a little bit of it but no more.
    - *d.* Tell her how horrible it is.
    - *e.* Ask for something else.
    - *f.* Ask to leave.
    - *g.* Tell her to eat it.
    - *h.* Throw it on the floor.
    - *i.* Remarks—
2. When your mother says that she thinks you need to go to the bathroom, what do you do?
    - *a.* Go.
    - *b.* Go and try.
    - *c.* Say I don't need to.
    - *d.* Go to the bathroom but not do anything.
    - *e.* Leave the room where she is.
    - *f.* Do nothing.
    - *g.* Remarks—
3. When you're outside and your mother tells you to come in the house, what do you do?
    - *a.* Go in.
    - *b.* Say "no."
    - *c.* Say "wait a minute."
    - *d.* Ask to stay out longer.
    - *e.* Keep on playing.
    - *f.* Remarks—
4. When your mother tells you it's time to go to bed, what do you do?
    - *a.* Go to bed.
    - *b.* Ask if you can wait a while first.
    - *c.* I don't go to bed.
    - *d.* Go to your room and play.
    - *e.* Say you don't want to go to bed—it's early.
    - *f.* Remarks—
5. Suppose your mother tells you to let some other child play with a toy you are playing with, what do you do?
    - *a.* Say no, or say it's mine.
    - *b.* Give it to him.
    - *c.* Give it to him begrudgingly or with resentment.
    - *d.* Say after a while.
    - *e.* Say why you don't want that child to have it.
    - *f.* Let the child have it and then take the toy back.
    - *g.* Get something else for the child.
    - *h.* Remarks—
6. If your mother asks you to pick up your toys, what would you do?
    - *a.* Do it.
    - *b.* Do part of it.
    - *c.* Say that you can't.
    - *d.* Ask her to help you.
    - *e.* Tell her to do it.
    - *f.* Say that it's not your job.
    - *g.* Tell her that you didn't mess it up.
    - *h.* Say no.
    - *i.* Remarks—
7. Suppose you and your friends were playing on a swing and your friends said your turn was to go last, what would you do?
    - *a.* Go last.
    - *b.* Tell them I wanted to go earlier.
    - *c.* I'd play with something else.
    - *d.* Tell adult.
    - *e.* I wouldn't play with them.
    - *f.* I wouldn't go last.
    - *g.* I'd break the swing.
    - *h.* I'd hit them.
    - *i.* Remarks—
8. Suppose your mother says you can't play with someone you like, what do you do?
    - *a.* I do anyway.
    - *b.* Ask her why.
    - *c.* I wouldn't play with him.
    - *d.* I'd play with him when she wouldn't know.
    - *e.* I'd be angry with mother.
    - *f.* Cry.
    - *g.* Get another friend.
    - *h.* Remarks—
9. Suppose other children are playing and say you can't play with them, what would you do?
    - *a.* I wouldn't play with them.
    - *b.* Cry.
    - *c.* Play with someone else.
    - *d.* Tell an adult.
    - *e.* Play alone.
    - *f.* Ask why.
    - *g.* I'd offer them something.
    - *h.* Hit them.
    - *i.* Remarks—

Sympathy

1. Suppose you see a child fall down and hurt himself, what would you do?
a. Laugh.
b. Push him down again.
c. Ask him if it hurts.
d. Tell an adult.
e. Do nothing.
f. Help him.
g. Remarks—

2. Suppose you see a child get spanked, what do you do?
a. Laugh.
b. Say it served him right.
c. Ask him why he got spanked.
d. Feel very sorry that he's spanked.
e. Avoid the person doing the spanking.
f. Avoid watching or seeing the situation.
g. I don't know; I wouldn't do anything.
h. Remarks—

3. If you saw some children hitting a boy, what would you do?
a. Hit him too.
b. Laugh.
c. Watch the other children hit him.
d. Ask why they're hitting him.
e. Tell them not to hit him.
f. Hit the children.
g. Run away.
h. Tell an adult.
i. Remarks—

4. If you saw a little child who was crying, what would you do?
a. Pay no attention to him.
b. Tell him to go home.
c. Ask him what the trouble was.
d. Tell an adult.
e. Help the child.
f. I don't know; I wouldn't do anything.
g. Remarks—

5. If you had some candy and your friend didn't, what would you do?
a. Give him some of mine.
b. Tell him that he can't have any of mine.
c. Tell him to go get some for himself.
d. Tell him that he shouldn't have candy.
e. Get some for him.
f. Do nothing.
g. Don't let him know you have candy.
h. Remarks—

6. If a child breaks his toy, what do you do?
a. Laugh.
b. Don't let him play with my toys.
c. Let him play with one of your toys.
d. Tell an adult.
e. Tell him not to worry, that he'll get a new one.
f. Do nothing.
g. Try to fix it or buy him another.
h. Remarks—

7. If your mother was sick and went to bed, what would you do?
a. Laugh.
b. Tell her to get up.
c. Pay no attention to her or play.
d. Tell her it serves her right.
e. Ask her if it hurts.
f. Ask her if you could help her.
g. Tell an adult.
h. Assist, or be quiet.
i. I'd take care of myself.
j. Remarks—

8. Suppose you saw a little bird that was hurt and couldn't fly, what would you do?
a. Leave it. Do nothing. I don't know.
b. Tell an adult.
c. Take it home.
d. Kill it.
e. Give it physical care.
f. Curious.
g. Remarks—

Strictness of Training

A. Cleanliness

1. Suppose one of your friends said "let's make mud pies," what would you do?
a. Join him.
b. Join him, but with caution.
c. Tell him it's too dirty.
d. Say no.
e. Ask my mother.
f. Watch play in the mud pies.
g. Remarks—

2. Suppose you were eating ice cream and it dripped on your clothes, what would you do?
   - *a.* Change clothes.
   - *b.* Clean it off.
   - *c.* Tell an adult.
   - *d.* Keep on eating.
   - *e.* Ask for more ice cream.
   - *f.* Look at it.
   - *g.* Remarks—

3. Suppose you were playing ball and your ball fell in a garbage can, what would you do?
   - *a.* Get it out.
   - *b.* Tell an adult.
   - *c.* Get another ball.
   - *d.* Stop playing.
   - *e.* Leave it there.
   - *f.* Get it out with device or covering.
   - *g.* Remarks—

B. Sex

1. Suppose you saw a little girl pulling up her dress, what would you do?
   - *a.* Tell her to pull it down.
   - *b.* Tell an adult.
   - *c.* Cry "Shame on you."
   - *d.* Laugh.
   - *e.* Look at her.
   - *f.* Avoid her.
   - *g.* Ask her what she's doing.
   - *h.* Nothing. I don't know.
   - *i.* Tell her it isn't right.
   - *j.* Remarks—

2. Suppose a little boy pulled down his pants in front of you, what would you do?
   - *a.* Tell him to pull them up.
   - *b.* Tell an adult.
   - *c.* Tell him it isn't right.
   - *d.* Say "Shame on you."
   - *e.* Laugh.
   - *f.* Look at him.
   - *g.* Avoid him.
   - *h.* Ask him what he is doing.
   - *i.* Nothing. I don't know.
   - *j.* Hit him.
   - *k.* Remarks—

C. Toilet

1. Suppose a little boy or girl wet their pants, what would you do?
   - *a.* Laugh.
   - *b.* Say "Shame on you."
   - *c.* Tell adult.
   - *d.* Feel sorry.
   - *e.* Say you shouldn't do that.
   - *f.* Avoid him.
   - *g.* Nothing. I don't know.
   - *h.* Remarks—

D. Nudity

1. Suppose you wanted to ask your mother (or father) something and she was taking a bath, what would you do?
   - *a.* Ask her.
   - *b.* Ask her later.
   - *c.* Get in tub too.
   - *d.* Help her.
   - *e.* Go away.
   - *f.* Ask permission to enter.
   - *g.* Remarks—

Ethics

1. Suppose you broke something of your mother's that she liked a lot, what would you do?
   - *a.* Tell her.
   - *b.* Replace it or fix it.
   - *c.* Hide pieces.
   - *d.* Hide self.
   - *e.* Cry or feel bad.
   - *f.* Not tell her.
   - *g.* Solicit aid.
   - *h.* Apologize.
   - *i.* Remarks—

2. Suppose you wanted a nickel and you saw your mother's pocketbook lying on the table, what would you do?
   - *a.* Leave it there.
   - *b.* Take it.
   - *c.* Take it and tell her.
   - *d.* Ask for money.
   - *e.* I'd use my own money.
   - *f.* Remarks—

3. Suppose you found a ball that you knew belonged to a friend of yours, what would you do?
   - *a.* Return it.
   - *b.* Keep it.
   - *c.* Hide it.
   - *d.* Tell an adult.
   - *e.* Tell the friend.
   - *f.* Leave it there.
   - *g.* Remarks—

4. Suppose you were playing ball and broke a window, what would you do?
   - *a.* Run away.
   - *b.* Tell an adult.
   - *c.* Not tell an adult.
   - *d.* Apologize.
   - *e.* Replace it.
   - *f.* Remarks—

5. Suppose you ate part of your dinner. There was some left but you aren't hungry any more, what would you do?
   - *a.* Eat it anyway.
   - *b.* Tell an adult I'm not hungry.
   - *c.* Leave it.
   - *d.* Give it to pet.
   - *e.* Leave table.
   - *f.* Cry.
   - *g.* Hide it.
   - *h.* Remarks—

Miscellaneous

1. What do you think children like you hate the most?
   - *a.* Dark.
   - *b.* Animals and insects.
   - *c.* Specific playmates.
   - *d.* Children.
   - *e.* Toy.
   - *f.* Food.
   - *g.* Father.
   - *h.* Mother.
   - *i.* Me.
   - *j.* Remarks—

2. What do you think children like you are afraid of?
   - *a.* Supernatural beings.
   - *b.* Older child.
   - *c.* Automobiles.
   - *d.* Being bad.
   - *e.* Being kidnapped.
   - *f.* Getting lost.
   - *g.* Animals and insects.
   - *h.* Punitive agents, police, gypsies.
   - *i.* Me.
   - *j.* Remarks—

3. Suppose you could be a baby again, how would you feel?
   - *a.* I wouldn't like it.
   - *b.* Mother would take care of me.
   - *c.* I'd be in a crib all day.
   - *d.* I'd drink milk, eat different foods.
   - *e.* I'd feel sick.
   - *f.* I'd like it.
   - *g.* Remarks—

4. Suppose you could be a grown-up right away, how would you feel?
   - *a.* I'd be glad.
   - *b.* I wouldn't want to because of school or other activities first.
   - *c.* I'd have children.
   - *d.* I'd be [some adult occupation].
   - *e.* Rejection of adult responsibility.
   - *f.* I'd rather be small.
   - *g.* I'd have adult authority and independence.
   - *h.* I'd have adult possessions.
   - *i.* Remarks—

5. What person do you think you are most like?
   - *a.* Mother.
   - *b.* Father.
   - *c.* Sibling.
   - *d.* Cowboy hero.
   - *e.* Friend.
   - *f.* Another adult.
   - *g.* Myself.
   - *h.* Remarks—

6. Suppose you could have a wish come true, what would you wish?
   - *a.* Food.
   - *b.* Toys, clothes, pets, etc.
   - *c.* Siblings.
   - *d.* More affection and security.
   - *e.* Less punishment.
   - *f.* Good time.
   - *g.* Didn't have to go to school.
   - *h.* Alternate sex.
   - *i.* Athletic hero.
   - *j.* Remarks—

APPENDIX B

CHILD-REARING PRACTICES INTERVIEW WITH THE MOTHER

Presentation of the Interviewer: Ask the questions at the top of each of the following sections. Try to maintain a spontaneous flow of description and remarks from the mother. Check answers to any of the questions as relevant data appear in her conversation. Do not ask specific questions and do not attempt to direct the mother's remarks by any show of preferential interest. Before each main topic of the interview is completed guide the conversation toward the questions she has not answered.

Presentation to the Mother: People aren't sure about what is best in bringing up children. We don't even know what mothers in general actually do in handling their children. As you know we are talking with a large number of mothers about this and we hope we can be more sure of these things when we are finished.

FEEDING

Take the problem of feeding, what happened with ————'s feeding and eating?

A. *Meeting Feeding Needs of Infant*

Yes No

1. Was infant awakened for bottle feedings?
2. Was hungry infant required to wait for a scheduled bottle feeding?
3. Was infant unable to take milk easily?

B. *Meeting Feeding Needs of Child*

1. Has child disliked any foods so that he won't eat them?
2. May child eat as much as he wants at mealtimes?
3. Does child eat between meals?
4. Has child shown a desire for a specific food to the exclusion of other foods?
5. Does child vomit easily?
6. Has child been markedly overweight?
7. Has child been markedly underweight?

C. *Nutritional Practices With Child*

1. Does child consume a quart of milk a day?
2. Does child have fruit or fruit juice once a day?
3. Does child have an egg every day?
4. Does child have a leafy vegetable once a day?
5. Does child have fresh meat daily?
6. Has child had Vitamin D (or sun lamp) regularly through winter season?
7. Has child had tonics or other supplementary vitamins?

D. *Emotional Factors at Feeding Time*

1. When the child doesn't eat something the parents usually:
   ———— *a.* Ignore it.
   ———— *b.* Encourage child to eat.
   ———— *c.* Coerce the child to eat.
2. Does child eat with other members of the family?
3. Has child cried in connection with mealtimes?
4. At what age was last bottle discontinued?
5. Did child want bottle at time it was discontinued?

INDEPENDENCE

We don't know enough about what five year olds can do for themselves and what are some of the things they don't do for themselves. Tell me about ————.

A. *Expression of Independence*

Yes No

1. Does child usually wash his face (except for atypical circumstances such as fatigue, special occasions, etc.)?
2. Does child usually wash his hands?

3. Does child usually wash his teeth?
4. Does child usually bathe alone?
5. Does child usually wash his hair?
6. Does child usually dress himself?
7. Does child play outdoors without adult supervision?
8. Does child perform chores or help around the house?
9. Does child refuse help from parents for things he knows how to do?

B. *Expression of Dependence*

1. Does child often request help for a task he has formerly accomplished or probably can do alone?
2. Does child without obvious provocation or difficulty often demand attention or affection (by whining or wanting to be embraced and babied)?
3. Does child go to parents for comfort or support because of minor accident or difficulty (other than quarrels)?
4. Does child go to parents about quarrels with children?
5. When the child has a quarrel with a playmate, does the parent intervene?

C. *Severity of Training For Independence*

1. When the child is in pain:
   ________ *a.* He is soothed or reassured.
   ________ *b.* His pain is ignored.
   ________ *c.* He is reproved for the expression of pain.
2. When the child is lonely or afraid:
   ________ *a.* He is soothed or reassured.
   ________ *b.* His feelings are ignored.
   ________ *c.* He is reproved for the expression of fear or loneliness.
3. When the child makes errors in tasks he has done:
   ________ *a.* He is reassured or aided.
   ________ *b.* His errors are ignored.
   ________ *c.* He is reproved for his errors.

D. *Safety Training*

________ *a.* Child is repeatedly warned to avoid harmful situations.
________ *b.* Child is occasionally taught to avoid harmful situations.
________ *c.* Child is not taught to avoid many harmful situations.

TOILET TRAINING

Another part of bringing up children is training them to use the bathroom. Tell me about ________.

A. *Initiation of Training*

Yes No

1. Was the time for the child's bowel movements predictable?
2. At what age was child first placed on the toilet?
3. How many times a day was the child put there?
4. How long did the child remain?
5. Did the child use the toilet during the early placements?
6. Did child scream or cry at the beginning of training?
7. After training was started did the mother stop and begin again later?

B. *Severity of Training*

1. At what age did child have all bowel movements in the bathroom?
2. At what age was child usually dry all day?
3. At what age could child usually go through the night without wetting or being picked up?
4. At what age did child go to inappropriate part of house for bowel movements or urination?
5. Does child ever sporadically wet during the day?
6. Does child ever sporadically wet during the night?
7. Was child scolded for soiling?

8. Was child ridiculed for soiling?
9. Was child hit for soiling?
10. Was child talked to constructively about soiling?
11. Did child have to wear soiled clothes as punishment?
12. Did child have to smell or contact soiled clothes?
13. Has child often gone more than one day without a bowel movement?
14. Has child had diarrhea?
15. Has child cried or been fearful after soiling during the day?
16. Has child cried or been fearful after soiling at night?
17. Does an adult reassure, quiet or comfort child after soiling?

C. *Interest of Mother*

1. During early training did the mother stay with the child in the bathroom?
2. Does the mother now stay with child?
3. Does the mother look at the child's stool?
4. Does the child routinely tell the mother when he has had a bowel movement?
5. Were child's messes disgusting to the mother?

AGGRESSION

We believe that it's perfectly natural for young children to get mad easily. What sort of situations seem to make ______ angry? How does it work out? What do you and your husband try to do?

A. *Expression of Aggression*

Yes No

Mother

When he is angry at the mother:

1. Does he try to strike her?
2. Does he scold her, call her unpleasant names?
3. Does the child leave the room, sulk, or become silent?
4. Does he cry?
5. Does he hold his breath, scream, roll on the floor, etc.?
6. Does he break objects or do other spiteful things?
7. Does he try to control himself?

When the child is angry with the mother does the mother:

8. Give in to the child's wishes?
9. Separate herself from the child?
10. Try to divert the child's interest?

When the mother scolds or punishes him:

11. Does he try to defend himself by explaining his circumstances?
12. Does he try to defend himself by counter-accusations?
13. Does he spontaneously promise better behavior?
14. Does he apologize?
15. Does he offer or agree to restitution or compensation?
16. Does he involve his father?

Father

When he is angry at the father:

1. Does he try to strike him?
2. Does he scold him and call him unpleasant names?
3. Does he leave the room, sulk, or become silent?
4. Does he hold his breath, scream, roll on the floor, etc.?
5. Does he cry?
6. Does he break objects or do other spiteful things.
7. Does he try to control himself?

When the child is angry at his father does the father:

8. Give in to the child's wishes?
9. Separate himself from the child?
10. Try to divert the child's interests?

When the father scolds or punishes him:

11. Does he try to defend himself by explaining his circumstances?
12. Does he try to defend himself by counter-accusations?
13. Does he spontaneously promise better behavior?
14. Does he apologize?
15. Does he offer or agree to restitution or compensation?
16. Does he involve the mother?

Yes No

Children

When he is angry at other children:

1. Does he strike them?
2. Does he scold them and call them unpleasant names?
3. Does he cry?
4. Does he leave, sulk, or become silent?
5. Does he scream, spit, or lose control?
6. Does he break objects or do other spiteful things?
7. Does he try to control himself?

B. *Severity of Training*

1. Is child usually spanked or hit for being aggressive toward the mother?
2. Is child usually scolded or reprimanded for being aggressive toward the mother?
3. Is child deprived of some privilege for being aggressive toward the mother?
4. Is child usually spanked or hit for being aggressive toward the father?
5. Is child usually scolded or reprimanded for being aggressive toward the father?
6. Is child deprived of some privilege for being aggressive toward the father?
7. Is child usually spanked or hit for being aggressive toward children?
8. Is child usually scolded or reprimanded for being aggressive toward children?
9. Is child deprived of some privilege for being aggressive toward children?
10. Do parents explain, interpret or reason with child about aggressive behavior?
11. Do parents ever encourage or compliment the child for aggressive behavior?
12. Does child rarely show anger toward mother?
13. Does child rarely show anger toward father?
14. Does child rarely show anger toward children?

CLEANLINESS

Some children don't seem to mind a little dirt and others don't like it at all but we don't know just how this goes. How is ________?

A. *Permissiveness*

Yes No

1. Were infant's or child's diapers changed immediately or very shortly after soiling?
2. Was child allowed to crawl on the ground or on a floor which might not have been clean at the time?
3. Did child play with feces?
4. Does the child go around indoors with soiled hands and clothes?
5. Does child remain clean for a prolonged period of time (all day Sunday, a long visit)?
6. When the child comes in from play, is he dirty?

B. *Severity of Training*

When the child gets dirty in dress or school clothes:

1. Is the child spanked?
2. Is the child scolded, reprimanded or ridiculed?
3. Does the mother clean him up without much comment?
4. Is he punished by having to wear the dirty clothes?
5. Is the activity prohibited?
6. Does the mother do nothing?

C. *Assumption of Responsibility*

1. Does child initiate washing procedures?
2. Does child help in cleaning up his room or toys?
3. Does child initiate a change to clean clothes?

D. *Cleanliness Habits*

1. The child is bathed
   ________ *a.* every day
   ________ *b.* two to three times a week
   ________ *c.* one time a week or less.
2. The child's under clothes are changed
   ________ *a.* every day
   ________ *b.* two to three times a week
   ________ *c.* one time a week or less.
3. The child's neck and ears are washed
   ________ *a.* every day
   ________ *b.* two to three times a week
   ________ *c.* one time a week.
4. At what age did child use a fork properly?

SLEEPING

What about sleeping?

A. *Isolation*

Yes No

1. Has child usually slept in his own room?
2. Has child usually slept in room with parents?

B. *Resistance to Sleep*

1. Does child usually fall asleep readily?
2. Does child awaken upon such slight provocation as small noises?
3. Has child walked in sleep?
4. Has child talked in sleep?
5. Does child have nightmares?
6. Has child required non-nutritive objects (except the thumb) to suck in order to fall asleep?

C. *Deviations*

1. Is child afraid of the dark or of being alone in bed?
2. Has child repeatedly rocked in bed or crib before going to sleep?

D. *Meeting Emotional Needs*

1. Has child had pets or toys in bed?
2. Does parent comfort the child when he awakens during the night?
3. Is child kissed or "tucked in" bed?
4. Has child had a night light left on in his room or the hall?

SEX TRAINING

Although everyone feels that we should know more about the sexual reactions and curiosity of children, no one has bothered to find out what parents have noticed in their children. Would you mind telling me about ————?

A. *Permitting Expression of Sexuality and Sex Curiosity*

Yes No

1. Does the child know that babies are carried by a woman?
2. Does the child know a male is required for conception?
3. Has child frequently observed unclothed adults?
4. Has child frequently observed unclothed infants and children?
5. Has child ever participated in sexual examinations or play with another child?

B. *Severity of Training*

1. Has child been told sexuality can hurt or injure him?
2. Has child been told sexuality is pleasurable but for adults or for people in love?
3. Has child been told sexuality is a bad thing?
4. Has child been physically punished for sex behavior?
5. Has child been reprimanded for sex behavior?
6. Has child been told to avoid companions or adults because of sex behavior?

Conclusions: I think we've covered some of the most common aspects of child rearing. Are there any additional matters that have seemed important to you in the past or at present that we haven't talked about?

Appendix C

CHECK LIST OF CHILD'S NERVOUS MANNERISMS AND FEARS
(Child-Rearing Practices Interview with the mother)

1. Nail-biting
2. Nose-picking
3. Hair-pulling
4. Tics
5. Thumb-sucking
6. Masturbation
7. Head-banging
8. Rubbing body parts
9. Rituals
10. Other
11. Fear of injury
12. Fear of water
13. Fear of fire
14. Fear of animals
15. Fear of doctors
16. Other fears

Appendix D

CONDITIONS OF CHILD REARING
(Continued from the Child-Rearing Practices Interview with the Mother)

1. *Physical Similarity*
   - *a.* One parent and child are strikingly similar in appearance. They have the same hair, eye and skin coloring and also have the same general features and physique.
   - *b.* One parent and the child are somewhat similar in appearance. There is about the same similarity and discrepancy in appearance as occurs among kin. There are no striking differences, however.
   - *c.* The child's appearance is strikingly different in some respect from that of both adoptive parents. Features, physique or coloring may be obviously different.
2. *Evidence of Interest in Child Care*
   - *a.* Full time boarding or caring for child by relative or friend.
   - *b.* Part time—day care, nursery school, baby-sitting.
   - *c.* None.
3. *Health of Adoptive Mother Since Adoption*
   - *a.* Mother has rarely been ill.
   - *b.* Mother has had illnesses, necessitating going to bed or missing work.
   - *c.* Mother has had one or more illnesses necessitating hospitalization.
   - *d.* Mother has died.
4. *Health of Adoptive Father Since Adoption*
   - *a.* Father has rarely been ill.
   - *b.* Father has had illnesses, necessitating going to bed or missing work.
   - *c.* Father has had one or more illnesses necessitating surgery.
   - *d.* Father has died.

5. *Parental Contacts*
   - *a.* Parents spend most evenings on leisure activities together.
   - *b.* Parents spend some of their evenings together and pursue some leisure time activities together.
   - *c.* Parents rarely spend evenings on leisure activities together.

6. *Mother-Child Contact*
   - *a.* Mother plans all her daytime activities so as to include the child. She usually does not work, has few activities outside the home, and spends most of the day with the child.
   - *b.* Mother plans some activities to include the child, but may work part time or be active in a few groups so that she is not with the child all of its day.
   - *c.* Mother plans most of her activities in a way which excludes the child. She may work all day or devote her time to pursuits outside the home. Except for details of child's routine, she is rarely with the child.

7. *Father-Child Contact*
   - *a.* Father plans all of his leisure time activities so as to include the child.
   - *b.* Father plans some of his activities so as to include the child but may have some other interests which exclude him.
   - *c.* Father plans his activities so that he spends almost all of his time away from the child.

8. *Community Service*
   - *a.* At least one parent is very active in community service. Parent may hold office in civic organizations or work repeatedly for community projects without pay.
   - *b.* At least one parent is moderately active in community service. Parent may attend civic organization meetings, contribute services to drives, and occasionally work for community projects.
   - *c.* Both parents are usually inactive in community affairs.

9. *Emotionality of Mother*
   - *a.* When angry, mother usually expresses her feelings without restraint. She may scream, throw things, or lose control in a tantrum.
   - *b.* When angry, mother expresses her feelings overtly but does not lose control.
   - *c.* When angry, mother rarely expresses her feelings directly.

10. *Emotionality of Father*
    - *a.* When angry, father usually expresses his feelings without restraint. He may scream, throw things, or lose control in a tantrum.
    - *b.* When angry, father expresses his feelings overtly but does not lose control.
    - *c.* When angry, father rarely expresses his feelings directly.

11. *Routine*
    - *a.* The home routine is coordinated around a fixed and inflexible schedule.
    - *b.* The home routine is organized around some fairly fixed schedule but there are minor variations.
    - *c.* The home routine is organized around a slight schedule so that there are extensive and frequent variations in routine from day to day.

12. *Religion*
    - *a.* There is consistent observance of all religious practices and prohibitions. The family is devout.
    - *b.* There is moderate observance of some religious practices and prohibitions. The family attends church at least occasionally but some departure from religious teaching is found.
    - *c.* Family is not interested in religion. They do not attend church and children are not educated in religion.

13. *Need Achievement*
    - *a.* One or both parents strive to do some things as well as possible and in some respect are recognized in the community or considered to be outstanding.
    - *b.* One or both parents strive to perform some things well but not outstandingly.
    - *c.* Both parents are interested in expanding only as much effort as is required for satisfactory performance.

14. *Expressed Disapproval of Parents*
    - *a.* Parents criticize and find fault with minute or unimportant details of child's behavior.
    - *b.* Parents criticize only as realistically warranted by child's behavior.
    - *c.* Parents rarely criticize and may tend to overlook or excuse shortcomings.
15. *Story Telling*
    - *a.* Parent and child usually read a child's story or picture book together.
    - *b.* Parent reads a story book to child. The child does not attempt to look at the written text o1 the pictures.
    - *c.* Parent rarely reads child stories.
16. *Conversation Between Mother-Child*
    - *a.* Whenever mother and child are together, conversation is almost continuous with both participating.
    - *b.* There is a fair amount of conversation between mother and child, but the child initiates or does most of the talking.
    - *c.* The mother is not talkative and tends to discourage the child from chattering.
17. *Films*
    - *a.* Parents exclusively attend movies which have been acclaimed for their artistic or dramatic value.
    - *b.* Parents occasionally attend movies which have been acclaimed for their artistic or dramatic value.
    - *c.* Parents rarely or never attend movies which have been acclaimed for their artistic or dramatic value.
18. *Radio (Politics)*
    - *a.* Parents exclusively listen to forums, round tables, and news analysts on the radio.
    - *b.* Parents occasionally listen to forums, round tables, and news analysts on the radio.
    - *c.* Parents rarely or never listen to forums, round tables, and news analysts on the radio.
19. *Culture (Music)*
    - *a.* Parents, though they may listen to other music also, frequently listen to opera or classical music.
    - *b.* Parents almost exclusively listen to semi-classical music.
    - *c.* Parents almost exclusively listen to popular music.
20. *Culture (Art)*
    - *a.* Parents regularly attend art displays or collect art objects.
    - *b.* Parents occasionally attend art displays or collect art objects.
    - *c.* Parents neither attend art displays nor collect art objects.
21. *Culture (Literature)*
    - *a.* Parents, though they may read other things also, read classics, outstanding novels, biographies, history, or literary magazines.
    - *b.* Parents almost exclusively read book club selections and popular magazines.
    - *c.* Parents almost exclusively read mysteries, comics, westerns, adventures, and pulp magazines.
    - *d.* Parents read only the newspaper for recreation.

22 *Culture*
- *a.* Parents frequently play musical instruments, paint, or sculpture.
- *b.* Parents occasionally play musical instruments, paint, or sculpture.
- *c.* Parents never play musical instruments, paint, or sculpture.

23. *Recreation*

Indicate how frequently one or both parents engage in the following:

Frequently or Regularly	Occasionally	Rarely or Never	
			1. Movies
			2. Radio
			3. Television
			4. Traveling away from home on pleasure trips
			5. Legitimate theater
			6. Cards, checkers, chess
			7. Pool and billiards
			8. Gardening
			9. Camping, fishing, hunting
			10. Dancing
			11. Collections of (non-art) objects
			12. Sailing and motor boating
			13. Golf
			14. Tennis, badminton, squash, handball
			15. Lectures
			16. Reading
			17. Playing football, baseball
			18. Horse races
			19. Other:

24. *Mother's Original Educational and Vocational Aspirations*
    - *a.* Mother's original educational and vocational aspirations are markedly superior to her present life situation.
    - *b.* Mother's original educational and vocational aspirations are slightly superior to her present life situation.
    - *c.* Mother's original educational and vocational aspirations are in line with her present life situation.

25. *Father's Original Educational and Vocational Aspirations*
    - *a.* Father's original educational and vocational aspirations are markedly superior to his present life situation.
    - *b.* Father's original educational and vocational aspirations are slightly superior to his present life situation.
    - *c.* Father's original educational and vocational aspirations are in line with his present life situation.

26. *Family Incompatibility*
    - *a.* The parents have contemplated divorce or trial separation.
    - *b.* There is evidence of deep incompatibility between parents but not sufficient to lead to a breakdown of the home.
    - *c.* There are minor differences of interest and opinion between parents but they are basically compatible.
    - *d.* The parents are harmonious and compatible.

Appendix E

MULTIPLE RATING SCALE

(A copy of this form is to be rated independently by both the psychologist and social worker)

1. *Acceptance of Child*
   - *a.* Parents completely accept the adoptive child as their own. They note with pleasure physical or behavioral similarities between the child and members of the adoptive family.
   - *b.* Parents do not wholeheartedly accept the adoptive child as their own. They may observe a few faults of the child or respects with which he compares unfavorably with family.
   - *c.* Parents do not accept the adoptive child as their own. They regret the adoption and would, at one time, have liked to have annulled it if possible.

2. *Solicitiousness Towards the Child*
   - *a.* Mother watches or fusses over child almost constantly. Is worried frequently and exaggerates the significance of minor ailments or difficulties.
   - *b.* Mother is concerned about the child's safety and health but trusts the child alone and allows him to take ordinary risks.
   - *c.* Mother is so unconcerned in major matters about the child's safety and health as to appear neglectful and irresponsible.

3. *Affectionateness of Mother*
   - *a.* Mother is overly affectionate with child. She kisses and fondles the child so frequently during the day that she probably interrupts the child's activities.
   - *b.* Mother is frequently affectionate with child. She will fondle and kiss him often but not to the extent of interfering with the child's activities.
   - *c.* Mother rarely fondles or kisses child.

4. *Accelerational Attempt*
   - *a.* Parent gives special training to child in order to assure his superiority with respect to other children.
   - *b.* Parent gives special training to child in order to assure his equality with other children.
   - *c.* Parent gives no special training to child in order to assure his equality with other children.
   - *d.* Parent exerts special measures to maintain immaturity of child by comparisons with other children.

5. *Flexibility of Parents' Expectations*
   - *a.* Parents have fixed and rigid ideas concerning child's education and vocation to which the child is expected to conform.
   - *b.* Parents have somewhat definite ideas concerning the child's education and vocation but do not insist that child conform.
   - *c.* Parents have no definite ideas concerning child's education and vocation.

6. *Mother's Enjoyment of Child*
   - *a.* The mother shows a great deal of real enjoyment and pleasure in most aspects of her child's behavior.
   - *b.* The mother shows some real enjoyment and pleasure in most aspects of her child's behavior.
   - *c.* The mother shows little real enjoyment and pleasure in most aspects of her child's behavior.
   - *d.* The mother shows no real enjoyment and pleasure in most aspects of her child's behavior.

7. *Mother's Reaction to Child*
   - *a.* Mother is critical of the child, describing his behavior in unfriendly and uncomplimentary ways.
   - *b.* The mother describes the child's behavior in an evaluative manner as if seeking to establish a value for herself.
   - *c.* The mother describes the child's behavior in an objective manner but accounts for it by favorable and protective explanations.
   - *d.* The mother is unable to recognize the child's behavior for what it is because of a marked protective bias.

APPENDIX F

RATING SCALE (HOME)

(Provided by the social worker)

1. *Physical Plant*
   - *a.* Home is in every way sufficient for meeting the child's physical needs with respect to warmth, quiet, cleanliness, toilet facility, ventilation, sunlight, and freedom from vermin and hazards.
   - *b.* Home is fairly sufficient for meeting physical needs (See *a*), but may be lacking in one or two respects.
   - *c.* Home is lacking in many requisites for meeting physical needs of child.

2. *Outdoor Play Space*
   - *a.* There is a large yard, lot, or park near the home where the child plays freely.
   - *b.* There is a small lawn, back yard, court, or play street near the home where the child plays freely.
   - *c.* There is little or no space outdoors where the child plays freely.

3. *Maintenance of Home*
   - *a.* The exterior of the home and any surrounding ground are very well maintained, and there is evidence that considerable time and care has been invested.
   - *b.* The exterior of the home and any surrounding ground are sufficiently well maintained that they appear tidy, but there is less time and care required for the upkeep.
   - *c.* The exterior of the home and any surrounding ground are not well maintained.

4. *Ownership of Home*
   - *a.* Parents own the house they live in.
   - *b.* Parents rent the house they live in.
   - *c.* Parents pay rent for a public type building, e.g. apartment house, hotel.

5. *Neatness*
   - *a.* Home is meticulous with everything in place. There are no misplaced toys or equipment either inside or outside.
   - *b.* The home is generally clean and neat but there is some disarray as if a few items had been recently in use.
   - *c.* The home is characterized by disarray. There are evidences that the home has not been recently cleaned.

6. *Stability of Residence*
   - *a.* One or both adoptive parents have lived in the present neighborhood or social group practically all their lives.
   - *b.* One or both adoptive parents have lived in the present neighborhood or social group ten years or more.
   - *c.* One or both adoptive parents have lived in the present neighborhood or social group five years or more.
   - *d.* Both parents are relatively new to the community.

7. *Contacts with Other Children*
   - *a.* Child has numerous playmates who are close in age (± one year) near the home.
   - *b.* Child has one or two playmates near the home or siblings of approximately his age.
   - *c.* Child has no other children or siblings of approximately his age near the home with whom he plays.

8. *Sociability (Visitors)*
   - *a.* The family has guests (including relatives and friends) frequently.
   - *b.* The family has guests (including relatives and friends) occasionally.
   - *c.* The family rarely has guests.

9. *Family Group Recreation*
   - *a.* Family enjoys group activities, e.g., they engage in family games, hobbies. They enjoy occasional excursions, picnics, carnivals, trips.

b. Family sometimes engages in group activities but these participations are not usually wholly pleasurable. One of the family members may become upset or lose interest.

c. Family enjoys group activities. They either do not engage in them or the activity frequently ends unpleasantly.

REFERENCES

1. Baldwin, A. L., Kalhorn, Joan & Breese, Fay H. Patterns of parent behavior. *Psychol. Monogr.*, 1945, 58, No. 3 (Whole No. 268).
2. Baldwin, A. L., Kalhorn, Joan & Breese, Fay H. The appraisal of parent behavior. *Psychol. Monogr.*, 1949, 63, No. 4 (Whole No. 299).
3. Baruch, Dorothy W., & Wilcox, J. Annie. A study of sex differences in preschool children's adjustment coexistent with interparental tensions. *J. genet. Psychol.*, 1944, 64, 281-303.
4. Bishop, Barbara M. Mother-child interaction and the social behavior of children. *Psychol. Monogr.*, 1951, 65, No. 11 (Whole No. 328).
5. Cummings, Jean D. The incidence of emotional symptoms in school children. *Brit. J. educ. Psychol.*, 1944, 14, 151-161.
6. Gesell, A. The measurement and prediction of mental growth. *Psychol. Rev.*, 1927, 34, 385-390.
7. Gesell, A. *Infancy and human growth.* New York: J. J. Little & Ives, 1928.
8. Gesell, A. The stability of mental growth careers. *Yearb, Nat. Soc. Stud. Educ., Part II*, 1940, 23, 149-160.
9. Gesell, A., & Amatruda, Catherine S. *Developmental diagnosis.* New York: Hoeber, 1941.
10. Gesell, A., Thompson, Helen, & Amatruda, Catherine S. *The psychology of early growth.* New York: Macmillan, 1938.
11. Hattwick, LaBerta A. Interrelations between the preschool child's behavior and certain factors in the home. *Child Develpm.*, 1936, 7, 200-226
12. Lafore, Gertrude G. Practices of parents in dealing with preschool children. *Child Develpm. Monogr.*, 1945, No. 31.
13. Lesser, G. Maternal attitudes and practices and the aggressive behavior of children. Unpublished doctor's dissertation. Yale Univer., 1952.
14. Meyer, Charlene T. The assertive behavior of children as related to parent behavior. *J. Home Economics*, 1947, 39, 77-80.
15. Miller, H., & Baruch, Dorothy W. Psychosomatic studies of children with allergic manifestations. I. Maternal rejection: a study of sixty-three cases. *Psychosom. Med.*, 1948, 10, 275-278.
16. Newell, H. W. A further study of maternal rejection. *Amer. J. Orthopsychiat.*, 1936, 6, 576-589.
17. Read, Katherine H. Parents' expressed attitudes and children's behavior. *J. consult. Psychol.*, 1945, 9, 95-100.
18. Symonds, P. M. *The psychology of parent-child relationships.* New York: Appleton-Century-Crofts, 1939.

(Accepted for early publication October 18, 1955.)

No. 409
1956

Vol. 70
No. 2

A Study of Adoptive Children: II. The Predictive Validity of the Yale Developmental Examination of Infant Behavior

By

J. R. Wittenborn and Others
Rutgers University

Vol. 70, No. 2 Whole No. 409, 1956

Psychological Monographs: General and Applied

A Study of Adoptive Children: II. The Predictive Validity of the Yale Developmental Examination of Infant Behavior[1]

J. R. Wittenborn
Rutgers University

With the collaboration of:

Myrtle A. Astrachan, Ph.D., Psychologist
Marjorie W. DeGooyer, M.S., Psychologist
W. Wallace Grant, M.D., Pediatrician
Irma Z. Janoff, Ph.D., Psychologist
Robert B. Kugel, M.D., Pediatrician
Barbara J. Myers, M.S., Psychologist
Annaliese Riess, Ph.D., Psychologist
Ellery C. Russell, M.S., Social Worker

INTRODUCTION

THE EXAMINATION developed at the Yale Clinic under the pioneering leadership of Dr. Arnold Gesell has become widely known and in many areas is used as the principal aid in the placement of adoptive children. Despite its wide acceptance and important role in many practical situations, the validity of this infant examination has never been explored by suitable investigation. Since Cattell (5),[2] Anderson (1), and Bayley (3) have been unable to find any evidence of useful validity for their infant examinations, some reasonable doubt concerning the validity of the Yale examination has arisen.

There are at least two possible interpretations of a general question concerning the validity of an infant examination. One interpretation concerns the question of whether or not the infant examination provides a relevant and acceptable evaluation of developmental status at the time of the infant examination. Gesell has commented on this aspect of validity in the following manner:

> Fundamentally the validity of the schedule here offered depends on the validity of the norms, the legitimacy of the category classi-

---

[1] The staff of the project wishes to express its appreciation to Dr. Donald Young, Director of the Russell Sage Foundation, and to Dr. Carl Hovland and Dr. Milton Senn of Yale University. The nature and the quality of this effort have been greatly influenced by such friends and advisers as Dr. Joseph Baldwin, Dr. Nancy Bayley, Dr. Harold Jones, Dr. Neal Miller, Dr. Robert Sears, and Dr. Helen Thompson. We are indebted also to Dr. Leonard Cottrell, Dr. Liselotte Fischer, Dr. Else Frenkel-Brunswik, Dr. Marjorie Honzik, Dr. Mary Cover Jones, Dr. Frederick Mosteller, and Dr. A. J. J. Van Woerkom. We are particularly grateful to the adoptive families who participated in this investigation, the child placement agencies who cooperated with us and guided our efforts, and to Adelaide Piscitelli, David Macklin, David Jenkins, and Peter Field who helped in the analysis of the data and the preparation of the manuscript.

[2] References for the present report are listed on page 114. The pagination of the present report is continuous with the first report of this series, entitled "A study of adoptive children: I. Interviews as a source of scores for children and their homes," *Psychol. Monogr.*, 1956, Vol. 70, No. 1 (Whole No. 409).

> fications, the appropriateness of each item for the category to which it is allocated, the soundness of the concept of maturity level, and the justness of using a sample of the child's behavior to indicate that level . . . (7, p. 219).

Our investigation is not concerned with validity of the infant examination as a basis for evaluating the developmental *status* of an infant.

Our investigation is concerned with a second aspect of validity. What are the predictive implications of the knowledge of the current developmental status of an infant provided by the infant examination? Gesell and his associates may also be quoted with respect to the aspect of validity:

> How consistent over a decade and more have been these mental-growth careers? In no instance did the course of growth prove whimsical or erratic. The behavior biographies give clear evidence of a high degree of latent predictability, even in infancy (6, p. 49).
>
> One of the greatest advances that we know of in adoption practice in the last twenty years has been the policy of giving careful developmental or psychological examinations to every single adoption candidate. Such examinations indicate whether an infant or child is of "normal" mentality, and therefore a good adoption candidate, or if he is so subnormal that he is not a fit candidate. They further can to some extent discriminate within the normal group children among average, above-average, and below-average endowment—thus giving a good clue as to what kind of home each child would best fit (10).

It is this predictive aspect of validity that provides the major question of the present investigation.

Our study of the predictive validity of the infant examination required measures for three classes of information. One class of information emerged from the infant examination; the second class of information described the child at the time of the follow-up study; the third class of information described some relevant facets of the child's environment.

It should be emphasized at the very beginning that a validity study is meaningful only to the extent that it shows evidence of relationship between the device to be validated and some criterion. Specifically, the most informative result of our investigation could emerge in those instances where our analyses reveal correlations between some aspect of the infant examination and some aspect of our criteria of development. Absence of relationship can be due to unreliability in the criterion, restrictions or inconsistencies in the sample, and numerous other factors, including limitations of the measuring instrument which is to be validated. Such limitations include lack of intrinsic validity for the criterion under consideration and lack of sufficient reliability to enable the instrument to be meaningfully related with any criterion.

Because of the ambiguous status of lack of relationships between the infant examination and criteria for development, we have sought to use all possible facets of the infant examination, all available measurements of childhood development, and all defensible subsamples of our adoptive children in an effort to show evidence of predictive validity. We may state this differently by saying that we have sought exhaustively to disprove the hypothesis of no predictive validity for the infant examination. As a matter of fact, this is the only defensible rationale in such an investigation because one can never prove an hypothesis of no relationship.

In the following paragraphs it will be noted that we have used several samples and subsamples in our exploration of the predictive possibilities of the infant examination. Since some of our adoptive children were placed in their adoptive homes selectively as a result of the infant examination (evidence for this will be provided later in this report), we have provided separate analyses for children who were placed in their adoptive homes *before* the infant examination and where presumably environmental differences resulting from selective placement could not be confounded with relationships between the infant examination and criteria for development.

Children who have been placed adoptively *after* the infant examination may not be ideal subjects for a validative study such as ours. Trends emerging for such children are suspect because they could occur as a consequence of confounding with environmental differences occurring as a result of selective placement. Such selective placement could result either from knowledge of true parents of the infants or from knowledge of the infants' standing on the infant examination. With respect to the danger of confounding as a result of knowledge of the true parents, we found a group of nonadoptive children living in their own homes; they provide a very desirable control. With respect to the possibility of confounding because of selective placement based on knowledge of the infant examination, we found certain measures of the environment to provide us with a desired control and means for analyses relevant to this particular selective placement hypothesis.

Because of numerous complexities introduced as a result of the practical nature of our investigation and as a result of the placement practices of social workers in southern Connecticut during the middle forties, some of our analyses are relatively easy to describe and others are relatively difficult to describe. Accordingly, we shall use simple subgroups in presenting the essential nature of our analyses. In our conclusions, we shall summarize the implications of our analyses for all the various measures and the various subgroups combined.

Since some of our discussions will be based upon a comparison of the relationships emerging from different subgroups, one might suspect that such differences are not due to the variables which we have in mind but are an artifact of some differences between subgroups such as range in test scores or such as age at the time the infant examination was administered, etc. It shall be apparent in our discussion of our subsamples that we have given proper consideration to these possibilities.

THE INFANT EXAMINATION

The infant examination developed by Gesell and his associates is not a psychometric test. It is a procedure for clinical examination and yields several kinds of descriptions and evaluations (7). Accordingly, any effort to evaluate the procedure must be concerned with its various facets and must attempt to validate them independently. In order best to describe the various facets, it would seem desirable first to describe the examination procedure in general. Our description will be brief, however, because authoritative publications are available which describe the examination fully, including the conditions under which the examination is given, the exact manner in which it is scored, and the precise nature of the norms employed (7, 8).

The infant examination procedure is based on observation of the infant's responses to a series of standard situations. These include response to the examiner, response to objects and situations, and in addition, certain reported responses contributed by the child's mother or nurse. The range of responses included in the examination procedure varies with the age of the child and with his developmental status:

it is unlikely that infants who differ from each other by several months in age are evaluated with respect to precisely the same kinds of responses. The infant examination procedure, like the Binet scales, changes in content for children of different ages. The opinion of many who employ the examination suggests that it may have its greatest value when administered to children whose ages lie between 12 and 40 weeks. (Some consider this to be due to the developmental pattern of infants; others consider it to be due to limitations in the content of the examination. Incidentally many practitioners claim that the examination may be used more reliably for identifying children who have a questionable developmental status than for distinguishing between children of superior developmental status.)

The examination yields part scores. These are available on the basis of an a priori classification of the items. This classification includes: fine motor, gross motor, linguistic, personal-social, and adaptive aspects of infant development. It is also common practice to combine the child's performance on these five parts of the examination procedure in order to provide a general estimate of infant development. It is customary to express the infant's developmental status with respect to developmental age equivalents. Such age equivalents are based on norms constructed from previous observations of children. It is possible, of course, to convert these developmental age equivalents to developmental quotients; this arithmetic adjustment for age, however, appears not to be a general clinical practice.

Although the examination provides a standard situation for observing the child, one finds diverse supplementary descriptions in the protocols. For example, in the examiner's written description of the child such words as mature, friendly, weak, etc., are often used. Some of these words occur with sufficient frequency in a set of protocols to justify their informal inclusion as an aspect or a part of the results of the examination procedure. In addition to such informal characterizations of the child and his responses, there are other descriptions that are more formal in their nature and have definite implications. For example, there are often diagnostic statements included with the examiner's description of the child, and some of these refer to the child's physiological or neurological status.

A principal feature of the examination is the general classification for the infant, a classification based upon a combination of the aspects of the child's development which are expressible formally in terms of age equivalent scores, the subjective impressions of the examiner, and various qualifying features which make the distinction between a clinical evaluation and a psychometric test. The classification of the infant was established with respect to levels, and included superior, high average, average, low average, dull, and sometimes such classifications as borderline and moron. It should be noted at the outset that in the case of many examinations no classification was given for the child. This ordinarily meant that for some reason or other the examiner was unwilling to commit himself with respect to the child's status, i.e., for some reason or other he considered the examination unsatisfactory or undependable. Sometimes this was due to some temporary indisposition on the part of the child. Occasionally, moreover, it occurred when the child's performance was definitely unsatisfactory and, if accepted at face value, would result in a very low or unfavorable classification. It is to the credit of the clinic and its individual examiners that they were extraordinarily cautious in classifying children whose performance was inferior. The procedure in questionable cases was to request further examination of the child.

The results of the examination were also expressed with respect to recommendations concerning the adoptability of the child (assuming that he was not already adopted). If, in the examination procedure, the child performed well (this usually was a basis for a confident examination), he was recommended as adoptable. Sometimes the recommendation was qualified and in some cases he was considered not adoptable. The recommendation of "not adoptable" was usually further qualified with the recommendation that the child be evaluated again at some later time. Other recommendations were made concerning the suitability of his current boarding or other temporary placement. For some cases it was suggested that the child was overplaced in the sense that the boarding home was too demanding or too stimulating for him; in other cases it was suggested that he was underplaced and that he could grow more rapidly in a stimulating home. It is difficult to determine in most cases how well the specific recommendations of the clinic were followed by the agencies or individuals requesting the examination.

It is probable that an examination would, in the case of many children, determine the immediate course of action (not necessarily the action recommended by the clinic, however). This seems particularly likely when one considers the conditions under which the infant test was given. An infant test could be initiated at the request of an adoptive parent, at the request of an agency handling the adoption, or at the recommendation of the court. Inasmuch as some one individual or a group of individuals had to take the initiative in securing an examination, it seems most likely that the results of such an examination would be used as an aid in planning the next step in the handling of the child. This possibility is anticipated in the introduction to the present report and provides one of the most obvious and also one of the most difficult complications in the present study.

It may be re-emphasized that the infant examination comprises many facets and that we should seek evidence for the validity of each of these various facets. They include the characterizations of the child, special diagnoses, developmental status with regard to the various parts of the examination and to the total examination; a general classification of the child based on performance on items and on general impression; recommendations with regard to adoptability, foster home, and other current circumstances. In some important respects the examination procedure was used cautiously; specifically, if a child did not perform well on the infant examination, subsequent examination was requested for him, classification may have been deferred, and presumably the possibility of immediate favorable placement was reduced. This means that concern was shown by the clinic for the possible untoward consequence of the spuriously low classification. It should be noted, however, that although the clinic itself tended to withhold classification until a satisfactory examination had been completed, it could not very well eliminate the possibility that a decision concerning the child would be made as a result of the first examination whether or not this examination were a satisfactory one from the standpoint of the clinic. If, for example, the outlook were unfavorable, this would be communicated to the agency and might prejudice their reaction to the child. At the very least, the qualified impression of the examiner would reduce the possibilities for a confident favorable placement of the child.

Since a second examination usually occurred as a consequence of a doubtful first examination, it is probable that children with more than one examination were classified more confidently on the final examination than they were on the first. It is not necessary, however, that a final examination should have a greater intrinsic validity than the first. On the contrary, there is some possibility that the results of the first examination could be prejudicial to the result of a later examination. Even if this possibility could rarely or never be demonstrably attributed to subjectivity among the examiners, it nevertheless seems quite possible that a qualified or unfavorable evaluation of the child at the time of the first examination could in many cases reduce the possibility of his having optimal opportunities for development. A qualified performance on the first examination could, in effect, perpetuate the kinds of conditions that existed at the time of the first examination so that the possibility of change in development on the part of the child could be restricted. Thus, it not only seems possible that the relationship between the final infant examination and some later criterion for development could be confounded by opinions based on the first examination, but that the relationship between the first examination and a final infant examination could also be confounded by a perpetuation of the same kind of factors that influenced performance on the first.

DESCRIPTION OF SUBJECTS

Since our sample of subjects was drawn from children who had been examined in infancy at the Yale Clinic of Child Development, many important properties of our sample are determined by the kinds of cases to which the infant examination had been applied at the clinic. Our survey of the records there indicated that during the period beginning in 1942 and ending in 1947 the children to whom the infant examination had been administered fell into three major groups (only a few were exceptional to this grouping).

One group comprised children who were ill, defective, or in some respect a cause of concern to their parents. Almost all of these children were living with their true parents, and none of them was being considered for adoptive placement. They were suffering from all sorts of disorders, many of them diagnosable as Mongolism, hydrocephaly, microcephaly, cerebral palsy, convulsions, birth injury, etc. Although the infant examination may have contributed to a refined or confirmed diagnosis for these children or to an exact evaluation of their current developmental status, this group of children was brought to the clinic for examination because of apparent or suspected pathology. As far as we know, none of the infants in this group was brought in for examination because at that time he was being considered for adoptive placement or because there was some question as to whether or not he should soon be considered for adoptive placement.

The second group of infants brought in during this period comprised those who were being considered for adoptive placement. The infant examinations were desired by child placement agencies, or by courts, or by adoptive parents. Although many of these children had a developmental status which was very modest if not questionable[3] and although some of them were weak and possibly ill, all of these children were being considered as adoptive candidates. It would appear to us that the children being considered for adoptive placement and the children being considered for clinical evaluation were quite generally and consistently different from each other so that the discernible defects that characterized the clinical group were relatively inconspicuous among the group being considered for adoption. Apparently very few grossly ill or defective children were being considered for adoptive placement during this period.

There was a third group of children examined during this period. They were living with their true parents and were examined as a part of the study of prospective members of the Yale Clinic Nursery School. This group was a relatively favored group; many of their parents were graduate students or members of the Yale faculty.

---

[3] It may be noted that for our younger sample the mean and standard deviation of the General Maturity Quotient on the infant examination were 96.60 and 15.77, respectively. Only one per cent of these children had General Maturity Quotients below 60, 15 per cent were 60 or above but below 80, 33 per cent were 80 or above but below 100, 46 per cent were 100 or above but below 120, and 5 per cent were 120 or above.

In our investigation, we drew our principal samples from the children who had been considered for adoptive placement. This means that very few, if any, grossly defective infants contribute to our results concerning the predictive validity of the Yale Developmental Examination. One should not generalize the results of our investigation to a possible use of the infant examination as a means of distinguishing between grossly defective children who are not adoptable by usual standards and questionably adoptable children whose development is slow or may have been retarded. Since the grossly defective or ill children did not appear in the group being considered for adoptive placement, one could suppose that for the most part these children are identifiable without the aid of the infant examination. Accordingly, as the infant examination was used in the Yale Clinic of Child Development during this period, its value as an aid in adoptive placement appeared not to rest on its value in making distinctions between the obviously ill or grossly defective and the infants of questionable development. We shall attempt to validate the use of the infant examination as an aid in the evaluation of infants being considered for adoptive placement.

Since the proper use of the infant examination may depend greatly upon the skill of the examiner, and since it is possible that, to an important degree at least, infant examinations are a standard medium for the skillful clinician to make his observations, it was decided that a study of the predictive validity of the examination should be based on the protocols of the *most approved* examiners.

Our sample of subjects is also influenced by the fact that we were interested in the predictive validity of the examination when applied to infants, particularly adoptive infants. As a consequence, we excluded from consideration all children who did not receive their first examination before they were 14 months of age.[4] Thus we are studying the predictive validity of an *infant* examination; we are not studying the predictive validity of an examination for very young children.

The records indicate that the examiners of the Yale Clinic of Child Development were very cautious in their evaluations and recommendations. It is amply evident that they sought to protect the adoptive children from a negative evaluation which could have been spuriously low or could have had spuriously negative implications. It would appear that they sought also to protect the adoptive parents from adoptive infants of questionable developmental potential. Because of these precautions, many of the infants were not promptly classified by the examiners. Instead, it was recommended that the infant be returned at some later time for further examination. It was further recommended that adoptive procedures be delayed until a more confident evaluation of the infant was possible. Accordingly, our samples could be seen as comprising children who had been confidently classified at the time of their first infant examination and children who had not, and were in turn subject to further examination. We made independent studies of the predictive implications of the infant examination for subgroups of children who differed with respect to the number of infant examinations they received. The various subgroups provided the same implications with respect to the question of the predictive validity of the infant examination. As could be imagined, however, the analyses for the children who had but one infant examination were the most simple; accordingly these are the analyses that receive a particular emphasis in the later sections. These analyses have the additional merit of being based on children whose first infant examination was considered to be most dependable.

---

[4] Much of our discussion of the infant examination will refer explicitly to subgroups who were between 12 and 40 weeks of age at the time of the infant examination.

One major characteristic of the sample was determined arbitrarily by us and had little to do with the testing practices of the clinic. This was the question of the age of the children at the time of our follow-up inquiry. In general, we wished to base our study on children who were as old as possible. In this way we could have not only knowledge of the children over a relatively long period of time, but we could also take advantage of the fact that there are more satisfactory measuring devices available for older children than there are for relatively young children. There were too few children in adolescence at the time of our investigation to enable us to make a satisfactory study of this stage of life. We believed that children in the first grade were in many cases going through a kind of turmoil or a period of uncertainty and rapid change, so that they would not be ideal children for a predictive study. This left us with two groups of children from which to choose; the five-year-olds who were old enough to test in a reliable manner, but not yet in the throes of the first grade, and the seven-, eight-, and nine-year-olds who were past the throes of the first grade but not yet involved in the preadolescent changes which assume such remarkably different patterns in different children. Sound arguments could be advanced favoring either the preschool group or the school group as the more appropriate for examining the predictive validity of infant examinations. Because of this, we decided to employ both samples, and we hope that the use of two samples of different ages will enable us to provide an element of replication in our study so that the implications for one sample can serve to confirm or to challenge the implications for the other sample. One hundred fourteen children were included in our younger sample, and 81 in our older sample.

We began our investigation by arranging a roster of adoptive children who had been examined before 14 months of age by approved examiners and who at the time of our follow-up inquiry would be at the proper age for either our younger or our older sample. In order to explore the nature and the degree of selective factors in our attempts to reach adoptive children for further study, we selected 170 names from our roster of younger children and sought intensively to include all of these 170 children in our study. The adoptive arrangements for 40% of these were handled independently and the adoptive arrangements for 60% of them were handled by a child service agency. It is interesting that we succeeded in including 56% of each of these groups in our study. (The balance of our younger sample of 114 and our older sample of 81 was secured in a similar way, except that accessibility became an important factor and relatively less effort was made to reach some of the cases who were distant, uncertain, or in some other respect presented complications.) It should be remarked that the families whose child had been placed independently were found to be cooperative with our study wherever we were able to reach them. The independent families who did not participate in our study were without known address, living in some distant part, or willing to disregard our inquiry.

Since we undertook our investigation in the expectation of providing research information that would be of value to child placement agencies, we sought to work closely with these child placement agencies, and tended to regard ourselves as their research staff. In the case of all agency placements, we called on the responsible agency, described our research plan to them, and asked them to help us to secure the interest and cooperation of the adoptive families. It is obvious that our agency placements were screened in some way or other by the agency. We cannot be sure how this may have affected the outcome of our study but it certainly did not prevent us from studying many families where the adoptive situation was uncomfortable either because of probable limitations in the home or in the child.

In any study such as the present one, one must be aware of the possibility that the children available for follow-up inquiry are in some important respect different from those not available, and as a consequence, the results of the investigation are in some way qualified and their general applicability limited. Because of this we made elaborate comparisons (with respect to infant records) between the younger children included and those not included in our follow-up study.

Our total roster of younger adoptive children who had been examined in infancy before 14 months of age and whose examinations were conducted by approved persons comprised 310 names. As indicated in an earlier paragraph, we made an intensive attempt to include a group of 170 of these children in our sample, and succeeded in including 95 of them. An additional 19 younger children were included on a "first come, first served" basis. Thus our study of the younger children is based on a total group of 114. Accordingly, our comparison between those available for a follow-up study and those not available for follow-up study involves the 114 younger children we had seen and the 196 younger children we had not seen.

With respect to the kinds of information that were descriptive of the infant, the results of our comparisons are reassuring. There was no reliable difference between the children we saw and the ones we did not see with regard to such factors as prematurity at birth, complications at birth, physical deviations, age at leaving the hospital, general maturity quotient on the infant examination, classification on the infant examination, outlook on the infant examination, examiner's recommendations concerning adoptability, and advice concerning the care of the child at time of infant examination. Furthermore, there were no differences with regard to such informal designations as active, attractive, attentive, sustained drive and interest, and stable. There were two such informal designations, *socially responsive* and *mature,* which offered some evidence of differences between the children in the follow-up group and those not in the follow-up group.[5]

---

[5] Since so few of the numerous comparisons show a trend, we may suspect that these trends arose by chance.

The reader may suspect that the children included in our study were different from those not included in our study with respect to the attainments or circumstances of the true parents. Although our data were not complete, we made comparisons between those seen and those not seen, with respect to such factors as the education and the occupation of the true father, and the education, occupation, and health history of the true mother. With regard to these classes of information, there was only one indication of a reliable difference; it was found that the true fathers of the children not included in our study had slightly, but reliably, higher educational attainments than the fathers of the children included in our follow-up study.

With respect to the kinds of early information which are not descriptive of the infant per se, but are descriptive of his care and general circumstances, we find that our comparisons yield some reliable trends. The children who went into an adoptive home after their first preliminary placement were somewhat more available to us than the others. This may suggest that adoptive parents who receive their children relatively early are more confident and cooperative about adoptive matters, including follow-up studies, than are parents who got their children after several preliminary placements. The examiner appears to have had some effect on the availability of the children for our study. We had learned in our interviews with the adoptive parents that in some cases the parents were in strong disagreement with the examiner and were displeased. Possibly displeasure arising from the examiner could have made some of our adoptive parents unwilling to cooperate in the follow-up study. This possible difference in the examiner cannot be shown to be due to intrinsic differences in the child (so far as we could discover, there was no practice in the Clinic that would involve giving all the "tough" cases to one examiner, etc.). Concerning the implications of differences among examiners, we have found that those examiners whose cases were underrepresented in our follow-up study tended in general to be examiners whose infants had general maturity quotients that were relatively low. Some examiners either drew relatively inferior children or were relatively pessimistic in evaluations of their children. Such differences among examiners are consistent from one test to another.

With respect to the data which are available to us, at least, there appear to be no important differences between children that we were able to include and those we were unable to include in our study. At least, any differences which our comparisons suggest cannot be seen to qualify the implications of our analysis.

Let us summarize our sampling procedure and provide a brief description of our samples. Our study is concerned with two major samples of adoptive children. There was one group of 114 *younger* children, most of whom were in their fifth or sixth year and none of whom was involved in the first grade. There was a second group of 81 *older* children. Most of these children were in their eighth or ninth year and none of them was in the first grade. Agency and independent adoptive placements were represented in both our younger and older sample, but not in equal proportions. Forty per cent of the younger sample (i.e., those more recently placed) were handled independently of any child placement agency, whereas 20% of the older sample were independent placements. Almost all of the children who had been placed by an agency were placed in their adoptive homes *after* they had received an infant examination, and we may suspect that the results of the infant examination may have been a factor in the agency's selection of the adoptive home. Almost none of the independent placements were placed in adoptive homes *after* they received an infant examination; accordingly, we have no reason for suspecting that the infant examination could have been generally important in the adoptive placement of the children who were handled independently.

Our report makes incidental use of a third sample of 35 children who live with their true parents and who were examined in anticipation of their attendance at the Clinic nursery school. This group was employed primarily as a control group in our study.

In the course of our validity analyses, it was necessary for us to make two kinds of distinctions among our children. First, we were interested in whether or not the child was placed in the adoptive home before the infant examination; this is relevant to the possibility of confounding. Second, we were interested in whether or not the first infant examination was optimal from the standpoint of possible validity; specifically, we were interested in whether or not the first examination permitted the examiner to classify the infant promptly and whether or not it was made at the best possible age to show validity. (Experts had advised us that the Yale infant examination may be most valid when it is given between the ages of twelve and forty weeks. It was indicated that infants under twelve weeks were too young to show a repertory of behavior that would permit valid evaluation, and it was indicated also that the growth norms for the items comprising the examination procedure were of such a nature as to make distinctions between older infants relatively inexact, certainly less exact than those possible for infants under forty weeks. This presumptive feature of the growth norms is congruent with the tables published by Gesell and Thompson [8].)

Table 1 provides a summary of our sample of adoptive children and a classification from the standpoint of relevant subsamples. It may be noted that 62 of our younger children received their first infant examinations at what was considered an optimal age and were classified as a result of the examination. Our division of these children into the 32 who were placed in their adoptive homes *before* the infant examination and the 30 who were placed in their adoptive homes *after* the infant examination is explicitly involved in one phase of our analysis and we have referred to these as the selected "before" and the selected "after" groups, respectively. Since we shall be involved in certain comparisons between these two groups, it should be noted that the means and the variabilities of these two groups are comparable with each other, not only with respect to age at the time of the first infant examination, but with respect to status on the first infant examination, later Binet IQ, and other relevant data.

CRITERIA OF DEVELOPMENT

There are many published reports based on the Yale Developmental Examination of Infant Behavior. Among them various statements may be found which imply that the examination procedure has a predictive significance for numerous different aspects of child development. So far as the writer is aware, however, there is no published statement

TABLE 1

A CLASSIFICATION OF THE CHILDREN WHO PARTICIPATED IN THE STUDY

Sample	The first infant examination led to an unqualified classification and was administered between the ages of 12 and 40 weeks*	The first infant examination did not result in a confident classification and/or examinations were administered after 40 weeks of age**	Total
I. THE YOUNGER ADOPTIVE SAMPLE:			
A. Infants placed in adoptive homes *before* the infant examination	32	22***	54
B. Infants placed in adoptive homes *after* the infant examination	30	30	60
Total	**62**	**52**	**114**
II. THE OLDER ADOPTIVE SAMPLE:			
A. Infants placed in adoptive homes *before* the infant examination	16	10	26
B. Infants placed in adoptive homes *after* the infant examination	37	18	55
Total	**53**	**28**	**81**
III. THE NURSERY SCHOOL SAMPLE (normal children living with their true parents and examined in infancy as part of plans for the Clinic Nursery School)			*35*

* Experts had advised us that the infant examination may be most valid when given in the 12 to 40 week range.
** Many of the children were not placed in adoptive homes until they had had two or more examinations.
*** Two of these were examined before 12 weeks; one at 9 weeks, the other at 11 weeks of age.

specifying the criteria appropriate for validating any of the possible predictive implications of the infant examination. This creates a difficult situation for one who seeks evidence for the predictive validity of the infant examination. If one were to be narrowly selective in deciding upon appropriate criteria, one would necessarily run the risk of omitting aspects of development which in the minds of some interested persons could be the basis for the most relevant of criteria. If, however, one wishes broadly to explore the predictive implications of the infant examination, two difficulties must be anticipated. One of these is the difficulty which is necessarily associated with gathering data relevant to numerous and diverse criteria of child development. The second difficulty emerges from the fact that standardized, generally applicable measures have not been developed for the personal and social development of children. For the most part, the devices available for studying the personal and social development of children are clinical procedures that are not only unvalidated but are of such a nature as to make validation most difficult.

Since we sought evidence for the predictive validity of the infant examination, we decided to measure and assess our children in all *feasible* respects. As we have indicated elsewhere, however, the fact that our children are adoptive children and that their availability to us is a result of the voluntary cooperation of their parents greatly qualifies what is *feasible*. Our examination procedure must be not only extraordinarily inoffensive and efficacious, but it must respect the tender status of adoptive children in our society.

In the following paragraphs we shall

describe the various criteria that were available to us. Some of the criteria, such as measures of intelligence, are tried and true; others, such as some of the scores emerging from the mother's description of the child are novel, untried, and without evidence of a high degree of reliability. With regard to the novel and untried criteria which we shall employ, we doubt if they should be depended upon to indicate the degree or precision with which such criteria might be predicted by the infant examination. Nevertheless, we believe that they possess sufficient reliability to satisfy a useful purpose. Specifically, if the infant examination has a predictive validity for the aspects of behavior represented by our novel criteria, we believe that the criteria we use will be sufficient to show this possibility.

In general, we hope to show evidence that the infant examination has predictive validity for the aspects of development which may be represented in our investigation. Since we use numerous criteria, it is not implied that we consider the infant examination committed to the prediction of all such criteria as these. In effect, we shall proceed to test a series of null hypotheses. Specifically, we shall proceed in the manner which is conventional to all modern investigations concerning the testing of hypothesized relationships between variables. We shall think in terms of a series of null hypotheses, all of which could be made explicit in such a form as "there is no relationship between [the specified aspect of the infant examination] and [the specified criterion for development]." We shall then proceed with our analyses in a manner calculated to refute as many as possible of the null hypotheses. It is probably unnecessary to remind the reader that a null hypothesis is a hypothesis to be *disproved,* and that it is illogical to presume to *prove* a null hypothesis or to establish its truth. Accordingly, our most meaningful results occur in those instances where we can refute a null hypothesis, because wherever we can refute it, the hypothesis that the infant examination has predictive validity is unchallenged and by implication could be true. Accordingly, wherever the null hypothesis is refuted, we may become interested in the degree of relationship between the aspect of the infant examination and the criterion of development that is revealed in our sample, and for our tried and true criteria at least, we may be interested in contemplating the predictive economy that could result from the use of an infant examination.

Failures to refute null hypotheses can be due to many causes, such as the use of criteria that have an inappropriate content, the use of criteria that have too little reliability to show a significant trend, the use of samples that are restricted or are in some other way inappropriate; and of course, it is possible that the predictor variable simply has no predictive value for any of the criteria under consideration.

There are doubtless many students of child development who would say, "The status of many features of child development is not clear; how can you arbitrarily specify what is good development and what is poor development in a child?" Because of the controversial status of some features of child development, we have eliminated them from consideration. Most of the aspects of child development which we have included in our criteria are ones with respect to which a large number of people would be in agreement, i.e., they would feel that they understood the relevance of the criteria for some of the values of our society. Our criteria may be outlined as follows:

I. *Intellectual Development*

A. The Stanford-Binet, 1937 Revision (12). The Binet test is considered by most people to be a good measure of the child's mental development, particularly with respect to those features of the mental development that represent the acquisition of verbal skills or that depend upon .employment of verbal skills for their expression.

B. The Arthur Point Performance Scale (2). A performance scale was included because it is possible that the acquisition of verbal skills is stimulated more by certain homes, whereas the acquisition of nonverbal abilities, e.g., spatial understanding and manipulating skills, is stimulated more by other kinds of homes. The Arthur was selected because it provides a comprehensive sample of important nonverbal abilities and because it has norms suitable for a relatively wide age range.

II. *Scholastic*

The possibility had been suggested that the developmental propensities measured best by the infant test may be brought out and made distinct by school experiences. It was decided, therefore, to employ two samples. One sample would comprise children who were old enough to be cooperative and responsive in a formal testing situation but not yet old enough to be in school; another sample would comprise children who were in school and had passed the throes of the first grade. Since it was decided that some of our children should be in school at the time of follow-up, a useful measure of their development would be some estimate of their scholastic achievements. We have used measures for reading, spelling, and arithmetic. The measures we employed (Gray Oral Reading Paragraphs Test, Stanford Achievement Tests Spelling List—Form D, and Stanford Achievement Test Primary Arithmetic Test—Form D) are standardized and widely accepted as appropriate evidences of achievement at the grammar school level (9). Data for scholastic achievement are, of course, not available for our younger subjects.

III. *Motor Development*

Motor development of our children was sampled by employment of two tests of strength—jumping and a hand dynamometer test. In addition, the proficiency with which children use their hands in the Arthur examination was rated. Ratings of alacrity and coordination were also made by the examining physician as he observed the child during the course of the pediatric examination. The strength tests were administered as part of the pediatric examination.

IV. *General Physical Development*

The general physical development of the children was appraised by collaborating pediatricians who were Fellows in the Yale Child Study Center. The doctors weighed the children; took their height; observed them during the course of a general pediatric examination which included an examination of the child's development, his sensory acuity, and his reflexive responses. The physical examination, although in general quite thorough, did not include an intensive study of any particular facet of the child's development. The examination was designed to reveal discernible, practically significant deviations in physical development of functions. The examination was not organized to reveal obscure or rare defects of structure or function.[6]

V. *Personal Development*

The aspects of the child's development most difficult to evaluate are those commonly designated as social or personal. Most of the devices that are available for evaluating a child's social or personal development have been prepared in clinical situations for the intensive evaluation of a given individual. They are not of such a nature as to permit the gathering of standard information for a large number of children. Because of this, the usual devices for studying children were not available to us, e.g., the Thematic Apperception Test or the Rorschach Test. It is only fair to point out that such well-known procedures are not only relatively idiosyncratic in their orientation but they themselves have not been validated and might indeed be more difficult to validate than the infant examination procedure per se. It seemed to us, therefore, that if it were necessary for us to use unvalidated criteria in order to examine certain aspects of the infant examination procedure, criteria should be adopted that have a certain amount of surface validity and that might possibly be subjected to independent validating studies. Evidence concerning the child's personal-social development is of three kinds:

A. An effort was made to sample the child's reaction to himself. In order to do this, we secured the children's responses to a series of hypothetical situations, for the most part situations familiar to them. We called this the Social Reaction Interview.

The information that we have gathered comprises many bits. It obviously would be undesirable to use each of these bits as one of our criteria. It seemed in the best interests of economy and clarity to combine or integrate these bits of information so that a conveniently limited number of variables of general meaning would be available.

The procedure for combining bits of information was very simple. All of the principal answers to each of the 111 questions in the Social Reaction Interview were correlated with the principal answer to every other question. Groups of questions which gave statistically consistent answers were formed. These groups were then refined by excluding any items that appeared to be psychologically inconsistent with the group. In this way groups or clusters of bits of information were formed that were both statistically and psychologically consistent. Any child could be given a score for a cluster simply by counting the number of answers he gave which fell in a particular cluster.

After such clusters were established, it was possible, for each of our children, to score the Social Reaction questions on the basis of each of these cluster scores. Such scores express the degree to which the kind of behavior implied by a cluster is characteristic of a child. Perhaps it is best to describe these scores as extensive scores rather than intensive scores. They do not imply how intense or severe any one response of a child might be, but instead indicate how general or pervasive a given kind of response might be for each child.

The manner in which this interview was used and scored is described in detail in *Interviews as a Source of Scores for Children and Their Homes* (See Part I of "A study of adoptive children"). Appendix A of that report includes a copy of the interview form which was used. The different scores which emerged from this interview are listed below:

---

[6] The form on which the results of this examination were recorded is included in the Appendix to this monograph.

	Name of Cluster	Split-half Reliability
Younger sample		
	Dependence on Adults	.88
	Aggression	.58
	Socialized Compliance	.70
	Takes Adult Role	.69
	Weakness-Avoidance	.57
	Constructive Approach	.58
Older sample		
	"Goody-Goody"	.59
	Responsible Attitude	.52
	Cooperation with Authority	.41

It will be noted that the procedure yielded more scorable clusters for the younger sample than for the older sample. We consider this to be reasonable because the increased self-awareness of older children introduces complexities which greatly obscure the value of such a direct, explicit question-and-answer procedure.

B. While the children were being examined, the mothers participated in a semistructured interview with one of the psychologists. The major portion of this interview had to do with child-rearing practices. The content of this interview was recorded in standard form and organized with respect to specific bits of information. Many of these bits of information were descriptive of the child as seen by the mother and provided the basis for a cluster-scoring procedure. This interview and the cluster-scoring procedure are described in *Interviews as a Source of Scores for Children and Their Homes* (See "A study of adoptive children: I"), and the standard forms we used for recording the interview are included in Appendix B of that report. The following list includes the cluster scores that were descriptive of the children:

	Name of Cluster	Split-half Reliability
Younger sample		
	Compulsive	.52
	Aggressive	.70
	Phobic	.57
	Dependent	.71
Older sample		
	Controlled	.78
	Spoiled Child	.78
	Dependent	.60
	Anxious-Aggressive	.80

It should be noted that the nature of the cluster scores appears to shift with the age of the children. This we consider to be consistent with generally observable changes in children of different ages and this is described in the report mentioned earlier in this paragraph.

C. During the course of the psychological and pediatric examinations the child was carefully observed and ratings of his behavior were made during that time. These ratings obviously are based on a restricted sample of behavior. Nevertheless, how a child behaves in response to a challenging situation is considered by many people to be relevant to his social and personal development. (Samples of the rating forms are included in Appendix A of the present report.)

It is obvious to the reader that we do not have any measure of the whole child and indeed we have not sought such measures; research for the most part is concerned with abstractions or generalizations and, in this particular sense, is different from clinical responsibility which must ever be concerned with the individual child as a total pattern of responses.

It will be seen that our criteria are diversified rather than intensive. We have not sought to examine the child exhaustively with respect to any particular aspect of behavior; we have sought instead to sample many of the different aspects of behavior which conceivably are relevant to the implications of an infant test, which are feasible in terms of the provisions of our study, and which are of some possible interest to those who are interested in the development of children.

In general, our criteria of development have been built to parallel the component parts of the infant examination procedure. It is suggested that the *adaptive* portion of the infant examination procedure might be most relevant to the measures of general intelligence provided by the Binet and the Arthur. The *fine motor* component of the infant-examination procedure might be relevant to the child's score on the Arthur, which is manual in implication, and with the psychologist's rating of the child's performance in manual respects. The *gross motor* portion of the infant examination procedure may seem to have a relevance for the gross motor ratings made by the pediatrician and the strength tests we have incorporated in the pediatric examination; in addition, this portion of the infant-examination procedure would conceivably be relevant to the structural development of the child as observed by the examining pediatrician. The *linguistic* aspects of the infant examination procedure would be relevant to the child's standing on the Stanford-Binet, which is a verbal measure of intelligence, the ratings of speech made during the examination, and by the (older) children's standings on reading tests and spelling tests. The *personal-social* aspect of the infant-examination procedure presumably would be relevant to a child's responses to the *Social Reaction Interview,* the manner in which he is described by his mother in the *Child-Rearing Practices Interview,* and the ratings of the child's responses in the testing situations made by the pediatrician and by the psychologists.

Perhaps these criteria would also be relevant to the informal characterizations of the child's personality which were made by the examiners and included in the informal description of the child.

Special neural or sensory evaluations based on the infant examination could be relevant to a child's over-all sensory competence as roughly estimated by the examining pediatrician, his neurological status as determined by the standard neurological examination, and by other details of the pediatrician's examination of a child's general development.

Thus it may be seen that we have sought in our small fraction of all conceivable criteria to sample kinds of behavior which would be relevant to the most conspicuous parts of the infant-examination procedure. In addition, the recommendations for the child with respect to adoptability are considered a part of the examination procedure. This, of course, can be validated in a sense by correlating it with the adoptive parent's reaction to the child.

It should be clear that a useful infant examination need not predict all these possible criteria. Any relationships found between the infant examination and the criteria could be an important reassurance to those who use an infant examination procedure. Nevertheless, if none of the aspects of the infant-examination procedure is found to have a useful degree of relationship with any of the criteria we have adopted, it would be very difficult to justify claims for the validity of the infant examination until criteria were devised which could be reliably predicted by some aspects of the infant examination. Specifically, if the infant examination does not predict any of the criteria we have adopted, then what of value does it predict? The criteria that we have sought to employ, though partly of our own devising, are not esoteric criteria; they include the kinds of behavior which both child psychologists and parents understand and consider to be important.

Table 2 provides a summary of the most promising of the various criteria which we employed in our exploration of the predictive validity of the various facets of the infant examination. Parts III, IV, and Section C of Part V of our outline of criteria will be seen to have a specific representation in Table 2. This representation requires comment. The physical examination in particular, and, to a lesser degree, the procedure for the psychological examination, provide opportunities for numerous ratings and evaluations in addition to specific measurements employed. Some of the informal features of these procedures are doubtless inconsistently defined and possibly unreliably observed. In order to

TABLE 2

CRITERIA OF DEVELOPMENT SELECTED AS MOST PROMISING ON THE BASIS OF PRELIMINARY EXPLORATION

I. Intellectual Development (younger and older samples both)
    A. The Stanford-Binet
    B. The Arthur Point Performance Scale

II. Scholastic (older sample only)
    Reading
    Spelling
    Arithmetic

III. Motor Development (younger sample only)
    Physical Examination Items
        Dynamometer, L.H.
        Dynamometer, R.H.

IV. General Physical Development (younger sample only)
    Physical Examination Items
        Height quotient
        Weight quotient

V. Personal-Social Development (younger and older samples both)
    A. Child's Reaction to Himself (Social Reaction Interview)
        Younger Sample:
        Dependence on Adults
        Aggression
        Socialized Compliance
        Takes Adult Role
        Weakness-Avoidance
        Constructive Approach
        Older Sample:
        "Goody-Goody"
        Responsible Attitude
        Cooperation with Authority
    B. Mother's Reaction to the Child (Child-Rearing Practices Interview)
        Younger Sample:
        Compulsive
        Aggressive
        Phobic
        Dependent
        Older Sample:
        Controlled
        Spoiled child
        Dependent
        Anxious-aggressive
    C. Ratings Made by Examiners
        Test Reaction Scale Items (younger and older samples both)
            Sensitivity to failure
            Reaction to failure
            Belligerence
        Physical Examination Items (younger sample only)
            Approach to examiner
            Reaction to undressing
            Activity
            Composite pediatrician's rating

eliminate the less meaningful data which these procedures provide and in order to select out the more promising data, all the various items, ratings, and bits of information were correlated with all of the facets of the infant examination and were correlated with our various measures of the environment. The elements of our data which produced no relationship of important degree or fewer evidences of relationship than could be expected on the basis of chance alone were discarded, and the remaining more promising elements are included in Table 2.

As we have noted in other paragraphs, our criteria for development vary greatly not only in content and source but in certain important technical respects. The criteria for mental ability, the Binet and the Arthur, are well known, well standardized, reliable, and valid in the sense that their relationships with other variables are known. Similar advantages may be claimed for the scholastic achievement tests. Some of the aspects of the physical examination, particularly the measures of size and strength, are well known and require little description or justification. Our measures for the social and personal aspects of children's development do not have all these advantages. Most of the cluster scores and the various ratings for personal and social development are novel and untried. Although we believe that little defense or justification for their purpose and content is required, it must be explicitly acknowledged that as measuring devices they are untried. We may suspect that as a group they are less reliable than our measures for mental ability. It is probable that some of these criteria are inappropriate for a study of the *degree* of relationship. Accordingly, we must not expect evidence of exact degree of relationship from these criteria. Much may be said in defense of these criteria, however. They meet certain practical requirements. As will be noted in our analyses, they are capable of yielding interesting and reliable correlations and accordingly, they obviously cannot be devoid of reliability. (In addition, we have shown promise of reliability for the cluster scores by the the use of the split-half method.) We believe, therefore, that if the infant examination has useful predictive validity for the behavioral content represented by these cluster scores, the cluster scores will be sufficient to show evidence of such validity. We find no reason for fearing that intrinsic validity for the content of these clusters will be obscured because of the imperfect nature of our measures.

Relationships Between Aspects of the Infant Examination and Criteria of Development

We sought evidence of relationship between all the various aspects of the infant examination and all of the criteria which were described in the preceding section. Our analyses involved certain intricacies and observed certain precautions that can be better illustrated than explained abstractly. Accordingly, we shall begin by a description of our study of the relationships between the General Maturity Quotient of the infant examination and Binet IQ at the time of our follow-up study. We begin with a discussion involving the General Maturity Quotient because this was the most promising aspect of the infant examination; almost all of the others showed no hint of validity.

It will be remembered that we expected to base our main analyses on the data for the younger group and to use the data for our older group to confirm or qualify the results of analyses for the younger group. It should be remembered also from our discussion of the infant-examination procedure and from our discussion of our samples that all of our children do not lend themselves equally well to a validity study such as ours. Specifically, some of the children were placed adoptively *before* an infant examination. It seems highly improbable for those children that any relationship between the infant examination and subsequent criteria for development could be confounded with environmental differences resulting from selective placement based on the infant examination. We know that there were other children, however, who were placed adoptively *after* the infant examination, and that for these children there was some chance that the infant examination could have been a factor in the selective adoptive placement of the children. As a matter of fact, during the period when our children were placed adoptively, it was common practice in many southern New England agencies to depend upon the infant examination as an important aid in making decisions concerning the selective placement of adoptive children. Accordingly, we could suspect that the relationships between the infant examination and the criteria could be confounded by differences in environment emerging or organized on the basis of differences in infant test scores. More may be said in this connection.

It will be remembered that the examiners were not equally confident of the examinations for all of the children, that the examiners refused to classify some children on the basis of the first infant examination and suggested that the child be returned at some later time for another examination which might permit a more confident classification of the infant. It would seem possible, therefore, that the infant examinations in which the examiner had the greatest confidence and which led to the prompt classification of the infants would be more likely to influence the selective placement of the infants than examinations which in some respects were qualified and which the examiner did not wholly trust. In this connection it should also be recalled that those who advised us indicated that, in general, the examiners had more confidence in this examination when given between the twelfth and fortieth weeks than when given at other ages. It is possible, therefore, that examinations given to infants during this optimal age period may have been more important in influencing the selective placement of infants than examinations given at other ages. Accordingly, our analyses will distinguish between those infants who received their first examination at an optimal age and were confidently classified and those infants who either did not receive their first infant examination at an optimal age or who were not confidently classified as a result of the first infant examination.

Table 3 illustrates the distinctions that we have made among our children. This table summarizes the results of our validity analysis where the predictor variable is the General Maturity Quotient on the infant examination and the criterion is Binet IQ at the time of follow-up. (Incidentally, Table 3 summarizes our analysis between the General Maturity Quotient and the IQ on the Arthur Point Performance Scale at the time of our follow-up, too.) It may be noted that Table 3 describes our results for both the younger and the older groups.

Let us first consider the results for the younger group. Here the younger group is divided into an "after" group,

TABLE 3

RELATIONSHIPS BETWEEN THE GENERAL MATURITY QUOTIENT ON THE INFANT EXAMINATION AND MEASURES OF MENTAL ABILITY AT THE TIME OF THE FOLLOW-UP STUDY

Sample	With Binet IQ		With Arthur IQ	
	r	*N*	*r*	*N*
I. THE YOUNGER ADOPTIVE SAMPLE:				
A. Infants placed in adoptive homes *before* the infant examination:				
1. The first infant examination led to an unqualified classification and was administered between the ages of 12 and 40 weeks	−.14	32	−.16	32
2. The first infant examination did not result in a confident classification and/or the examinations were administered after 40 weeks*	.15	21**	.13	21**
B. Infants placed in adoptive homes *after* the infant examination:				
1. The first infant examination led to an unqualified classification and was administered between the ages of 12 and 40 weeks	.55	30	.19	30
2. The first infant examination did not result in a confident classification and/or the examinations were administered after 40 weeks	.16	30	.03	30
II. THE OLDER ADOPTIVE SAMPLE:				
A. Infants placed in adoptive homes *before* the infant examination:				
1. The first infant examination led to an unqualified classification and was administered between the ages of 12 and 40 weeks	.20	16	−.06	16
2. The first infant examination did not result in a confident classification and/or the examinations were administered after 40 weeks	.42	7**	.62	7**
B. Infants placed in adoptive homes *after* the infant examination:				
1. The first infant examination led to an unqualified classification and was administered between the ages of 12 and 40 weeks	.15	37	.28	35**
2. The first infant examination did not result in a confident classification and/or the examinations were administered after 40 weeks	−.02	18	−.22	18
III. THE NURSERY SCHOOL SAMPLE	−.14	35		

Note.—The average correlation between the General Maturity Quotient and the Binet IQ for all the various adoptive subsamples except the one which provides evidence for selective placement (the 30 younger adopted children in the selected after group) is .09. If this average includes the nursery school group also, it becomes .05. If an average correlation is computed which includes all the various subsamples except the younger one where selective placement was indicated, the nursery school group, and the selected younger sample when environmental differences have been partialed out of the correlation one finds the average correlation between the General Maturity Quotient and the Binet IQ to be .08.

* Two of these were examined before 12 weeks; one at 9 weeks, the other at 11 weeks of age.

** We did not have complete data for all cases; therefore, some of these correlations are based on *N*'s slightly smaller than the total *N*'s in the subsamples (see Table 1).

the group where the infants were placed adoptively after their infant examinations. A "before" group was also indicated, this was a group where the infant was placed adoptively before the infant examination. In the "after" group one could suspect selective adoptive placement occurring in part as a result of the infant's standing on the infant examination. In the "before" group one could have no basis for suspecting that the infant examination could have been a factor in selective adoptive placement. It will be noted that both the "after" and the "before" groups are divided into two parts. One part which we have selected for special emphasis comprises those whose first infant examination was administered at an optimal age and led to a confident classification of the infant. The other part comprises those whose first infant examination either was not given at an optimal age or was in some manner qualified and did not lead to a confident classification.

We would anticipate that *selective* adoptive placement based on the infant examination would be particularly probable for the selected "after" group of children who received their infant examination at an optimal age and whose infant examination led to a confident classification. It is obvious that we must be careful in our use of these "after" groups in our validity analyses. It is possible that such groups could yield seriously misleading evidence incorrectly suggesting a relationship between the infant examination and some criterion, e.g., the Binet. Such a confounding of the relationship could emerge because of selective arrangement of the relevant environment occurring as a result of selective adoptive placement based on the infant examination.

In view of our concern about the possible effects of selective placement, the data provided by Table 3 are most interesting. The only reliable relationship between the General Maturity Quotient and the Binet IQ occurred for the selected "after" group where selective placement based on the infant examination was considered the most probable. Specifically, the only correlation for the younger children which might possibly be used to refute the hypothesis of no relationship between the General Maturity Quotient and Binet IQ is for the very subgroup where selective adoptive placement (and the confounding with the environment) seems most probable. It is interesting to look at the data provided by the older sample. None of its subgroups provides any evidence for the refutation of the hypothesis of no relationship. To be sure, one of the very small older subgroups produces a high correlation. This correlation is not reliably greater than zero, however, and scrutiny of our data shows that it is almost wholly the result of two cases.

If for the younger sample one were to exclude from consideration the subgroup where selective placement seemed most probable, the selected "after" group, the average correlation for the three remaining subgroups would be only .05. If one were to take the average correlation among the four subgroups comprising the older sample, a value of .14 would emerge. Thus it would appear that if one were to discard the subsample where selective placement would seem most probable, we would have no basis for refuting the null hypothesis.

Since the probability of selective placement is so great for the selected "after" group, we do not consider it to be appropriate for our analyses and it should be excluded from consideration. Because we suggest that this subsample be excluded from consideration and because of our interest in adoptive placement practices as they may have involved the infant examination, it is of interest to examine the plausibility of these selective placement hypotheses.

In order to show that selective placement could have occurred in the particular subgroup where we had suspected it, we conducted two parallel studies of possible relations between aspects of the infant examination and aspects of the child's environment. These parallel studies required two groups. One group comprised the 30 children who were tested at an optimal age, confidently classified on the basis of the first examination, and placed in adoptive homes *after* the first examination. The other group was also tested at an optimal age and confidently classified; they, however, (32 in number) were placed in adoptive homes *before* the first infant examination. Since the selected "after" group was for the most part agency placements and the selected "before" group was mostly independent placements, these two groups could have been somewhat different from each other with respect to infant-examination scores and measures for the environment. With respect to the infant examination, however, the differences between our selected "before" and selected "after" groups were not great. For example, in the selected "before" group (comprising 32 children) the mean and the standard deviation of the General Maturity Quotient on the first infant examination were 102.3 and 12.3, respectively. For the selected "after" group (comprising 30 children) the mean and the standard deviation on the first infant examination were 97.0 and 11.7, respectively. The difference between the means appears to be due to one extreme case. The difference between the standard deviations is not significant and accordingly we need not be concerned about differences in range of infant examination scores when we conduct parallel analyses for the selected "before" and selected "after" groups.

It is our particular interest to see whether or not there are some environmental differences which are predictable from the infant examination for the selected "after" group but not predictable from the infant examination for the selected "before" group. If the infant examination is used as a basis for selectively placing adoptive children in adoptive homes, we would expect correlations between some aspects of the infant examination and some aspects of the adoptive homes for groups of adoptive children who were placed by agencies *after* the infant examination. We would not expect such correlations to appear for groups of children who were placed adoptively *before* they received an infant examination.

Let us review some of the more interesting indications from this parallel analysis. Beginning with the *Ambition* cluster score for the home (see our report *Interviews as a Source of Scores for Children and Their Homes* for a discussion of environment scores), we find no evidence of important relationships for the "before" group. (The correlation between the General Maturity Quotient and the *Ambition* cluster score is -.17.) For the "after" group, however, we find indications of relationship between the General Maturity Quotient and the *Ambition* cluster score. (The respective correlation is .38.) Thus the "after" group not only provides a positive correlation but the correlation coefficients for the two groups are significantly different from each other.

There is no hint of important relationship between the *Harmonious* home cluster and the infant examination.

For the *No Ambition* cluster, however, there are indications of relationships. It is found that for the "after" group there is an inverse relationship between aspects of the infant examination and this cluster score; this suggests selective placement. Curiously enough, there is a positive relationship between the *No Ambition* cluster score and the respective parts of the infant examination for the "before" group.

Continuing with our discussion of the cluster scores, we find no indication of relationship between status on the infant examination and the *Rejection* cluster score for the home. Similar apparent lack of relationship characterizes our survey for the *Parental Incompatibility* cluster score and the *Overprotective* cluster score. The balance of our analyses for the various cluster scores suggests that selective placement did occur for those children who were placed after their infant examination. One interesting instance is the postitive relationship between good standing on the infant examination and a tendency for adoptive parents to punish aggressive expression in the children ("after" group).

For these two groups of children we find certain additional evidence for selective placement. For the selected "after" group there is indication that status on the infant examination is related with the education-occupation status[7] of the adoptive father ($r = .47$). There is no such indication for the selected "before" group ($r = -.12$). These correlations are reliably different.

With respect to the difference between the selected "before" group and the selected "after" group in apparent degree of correlation between the General Maturity Quotient on the infant examination and the Binet IQ at time of follow-up, certain additional comments may be offered. One might expect that the difference between these two correlations is somehow or other an artifact of some insignificant differences between the two groups with respect to age at infant examination. In order to examine this possibility minutely, it was possible to arrange a selected "before" group and a selected "after" group so that they were exactly matched with respect to age at time of infant examination. For these respective matched groups of 22 cases, the correlation between General Maturity Quotient and Binet IQ was again determined and the two correlations were found to be reliably different.

In order to pursue the hypothesis that the correlation of .55 for the *selected "after" group* was an artifact of selective placement based on the infant examination, it was decided that we should attempt to consider the effect of probable aspects of the agencies' selective placement practices on this correlation. In the preceding paragraphs we indicated that we have provided evidence of selective placement with respect to the Ambition–No Ambition[8] aspects of the adoptive home, and with respect to the education-occupation status of the home. Our conversations with the various agency representatives as well as the published reports of other investigators (4) indicate that the age of the adoptive parents and the duration of their marriage was also taken into consideration. Accordingly, we formed an Age–Duration of Marriage composite[9] score.

---

[7] The composite education-occupation scale forms a continuum representing a combination of educational and occupational achievement of the home, and was based on the following operation: If the adoptive father's occupation was business or professional or in the white-collar group, a score of plus one was arbitrarily assigned. If, however, the adoptive father's occupation was unskilled, semiskilled, or skilled, the weighting was minus one. Weighting of zero was given in all situations where we did not know the occupation of the adoptive father. The father's education was combined by allowing plus one if he had attended college, minus one if he had not attended college, and zero when we did not have this information. The mother's education was scored in a similar manner. These three sets of weights were combined algebraically to form a general composite score representing the education-occupation status of the home.

[8] This Ambition-No Ambition composite cluster score was formed by giving the score on the *Ambition* cluster a plus sign and the score on the *No Ambition* cluster a minus sign, and then taking the algebraic sum of the two scores.

If our selected "after" group had been very large indeed, it might have been possible for us to find subgroups which differed from each other with respect to the environmental factors resulting from selective placement, but which were homogeneous within themselves with respect to these environmental factors. An average correlation between the General Maturity Quotient and the Binet IQ for such subgroups of homogeneous environments would be a correlation free from the influence of differences in the environment resulting from selective placement. Our group is obviously too small to make possible such an analysis, but from the statistical standpoint the partial correlation coefficient is considered to be the equivalent of such an hypothetical average correlation.

Accordingly, a partial correlation between the General Maturity Quotient and the Binet IQ was computed for the selected "after" group. The partial correlation represents the correlation between the General Maturity Quotient and the Binet IQ after the influence of differences in the Ambition–No Ambition cluster score, the Age–Duration of Marriage score, and the Education-Occupation score have been eliminated from the original correlation. Under these circumstances the original correlation of .55 is reduced to .28. We consider this reduced correlation to be the best estimate of an average of correlations which might be found for hypothetical subgroups which differed from each other with respect to these environmental factors (seemingly organized on the basis of selective placement) but which within themselves were homogeneous in these environmental respects; we specifically regard this correlation as an estimate of an hypothetical correlation which is free from the aspects of confounding previously discussed.

Readers who prefer an hereditary explanation for individual differences may object to this partial correlation on the grounds that they consider both the differences in the infant examination and the differences in Binet IQ to be organized at least partly on the basis of common heredity and that any organization of the environment which occurred as a result of differences in the infant examination (or as a result of hereditary differences in the true parents) was in part at least an expression of the kind of hereditary factors which were responsible for the relationship between the General Maturity Quotient and the Binet IQ. Such readers could claim on the basis of their beliefs that our partial correlation analysis partialed out too much, i.e., it partialed out differences which were an expression of the relevant underlying heredity. In order to answer this question, we can offer two considerations. First, the correlation between the infant examination and the education-occupation status of the true parents is only .16. This provides no basis for believing that the infant examination is an expression of any prepotent hereditary factors of the true parents which are sufficient to result in differences in their educational-occupational attainments.

---

[9] The age-duration composite was prepared in the following manner: If the adoptive father were in his thirties, a score of plus one was arbitrarily given; if he were under thirty or forty or over, a score of minus one was given; if his age were unknown, a score of zero was given. The same scoring system was used for the age of the adoptive mother. In addition, if the couple had been married for seven or more years, a score of plus one was given; if they had been married for less than seven years, a score of minus one was given; and if the duration of their marriage was unknown to us, a score of zero was given. These three weights were added algebraically to give us a composite measure of the age of the adoptive parents and the duration of their marriage.

Our more effective answer to this question involves the use of a control group. We were able to include in our follow-up study 35 children living with their true parents. These children had been examined in infancy as a part of preparations for their enrollment in the Clinic nursery school. Unfortunately, the correlation for this group was −.14. If an arrangement where a congruence between characteristics of the true parents and characteristics of the child's home environment could produce a correlation between the infant examination and criteria of development, we should most expect such a correlation to emerge for this group. This discouraging correlation of −.14 obviously fails to provide any indication that our partial correlation procedure disregards an important hereditary component in the correlation of .55 for our selected "after" group (where selective placement based on knowledge of either the true parents or the infant examination could have occurred). Incidentally, this sample of 35 children provides an additional instance where an hypothesis of no predictive value could have been refuted (if refutable by the use of our criteria). Unfortunately, this sample provides no challenge to the null hypothesis (the hypothesis of no relationship).

It should be remembered that we sought evidence of relationship between each of the various facets, scores, and indices of the infant examination and each of our various criteria for development. The Classification on the infant examination, which includes both the quantitative aspects of the infant examination and certain qualitative aspects, was found to reveal a pattern of relationships with the Binet IQ which was very similar to the pattern revealed by the General Maturity Quotient. The selected "after" group showed a correlation between Classification and the Binet IQ of .60. The emergence of this particular correlation is quite suggestive of selective placement based on the infant examination. Incidentally, the correlation for the selected "before" group was −.06. Certain preliminary analyses showed that the examiner's descriptions at the time of the infant examination ("drive and interest") was reliably related with Classification but not with General Maturity Quotient.[10] We had found that the children who were characterized by "sustained drive and interest" tended to be placed in those homes which eventually showed relatively high scores on punishing for aggression. When differences in the Punishment for Aggression cluster score, as well as differences in Ambition–No Ambition, Age–Duration of Marriage, and Education-Occupation were partialed out of the correlation between Classification on the first infant examination and Binet IQ at time of follow-up, the correlation was reduced to .30. In this connection the reader may wonder whether infants who were characterized by sustained drive and interest developed into children who required punishment for aggression. In order to answer this question, we examined the relationship between "sustained drive and interest" and "punishment for aggression" for our selected "after" group and our selected "before" group. For our selected "after" group the relationship was significant, whereas for the selected "before" group the relationship was not significant. This is taken as a further indication that drive and interest were factors in the selective placement of the infants and not necessarily a factor which predetermined the kind of child-rearing practices that would emerge as a consequence of a child's "intrinsic" personality.

---

[10] This is consistent with the composite basis for the Classification as contrasted with the psychometric basis for the General Maturity Quotient.

The reader may be dissatisfied with our partial correlation analysis and we do not regard the results for the selected "after" group as first-rate evidence. This most obviously is a group where selective placement based on the infant examination is most probable and perhaps the evidence provided by this group should be excluded from consideration. It is of interest, however, to show that the indication of relationship between the infant examination and the Binet criterion of mental ability occurs for this group only and may be plausibly explained in terms of a selective placement hypothesis. Stated differently, by a consideration of the selective placement hypothesis, the indication for this most dubious subsample may be brought in line with the indications for all of the other subsamples.

As we terminate our discussion of the criteria for mental development, one may wonder why the selected "after" group with the strong evidence of selective placement based on the infant examination did not provide evidence of a correlation between the General Maturity Quotient and the Arthur IQ. The explanation for this may lie in the fact that the various environmental factors which the agencies seemingly considered in their selective placement practices are not shown by our analyses to be as consistently or as reliably related with the Arthur IQ as they are with the Binet IQ. From this one would infer that the differences in ability which the Binet represents and the differences in ability which the Arthur represents do not emerge in the same manner as a consequence of environmental differences. Our numerous analyses involving all facets of the infant examination and the criteria for mental development provide no basis for refuting the hypothesis that there is no relationship between the infant examination and the criteria for mental development. In view of all the relationships examined, fewer promising trends emerged than could be accounted for on the basis of chance alone, and these promising trends were found only for a subsample where selective placement based on the infant examination was not only anticipated on a priori grounds but was indicated on the basis of analyses. In addition, partial correlation analyses sustained and illuminated the selective placement hypothesis for the emergence of these few correlations. Despite the fact that we used numerous subgroups, two different age samples, and a sample of children living in their own homes, we could find no challenge to the hypothesis of no relationship, and the implications of our study are congruent with the implications of others who have examined this question (1, 3, 5).

Our data for the older sample included scholastic achievement criteria. There was no indication that any of these criteria was reliably related with any aspect of the infant examination.

When we consider certain "nonpsychological" aspects of the infant's development, we find some indication that the infant examination may have some predictive value. Specifically, it is found that the adaptive developmental quotient on the infant examination has a correlation of .45 with strength of grip in the left hand as measured by the dynamometer at the time of our follow-up examination. This correlation of .45 is found for the selected "before" group, a group in which we have no basis for expecting selective placement or confounding from any other source. Although this correlation is plausible in the sense that the adaptive developmental quotient involved the use of the hands, the correlation was not reliable for the other subgroups and could be an accident of sampling. Since the physical examination was not conducted with the older sample, we could not explore this relationship further.

The weight quotient (a weight/age ratio) showed an interesting correlation with the fine motor developmental quotient of the infant examination. The correlation for the selected "before" group was .44 and for the selected "after" group it was .20. The other correlations were less promising.

It is interesting to note that the most definite suggestions of predictive validity of the infant examination are the physical aspects of development. It may be remembered from our discussion of the infant examination that this was an aspect of development which received relatively little emphasis in the literature concerning the infant examination (an examination for infant behavior). We should not overgeneralize, however, and conclude that our data show definite evidence that the infant examination has predictive value for these and other physical aspects of development. Our data are suggestive and should be followed by further inquiry where the physical aspects of development are the focus of the investigation.

The various possible relationships between aspects of the infant examination and our various measures for the personal and social development of the children were examined for our different subgroups. The many correlations were small and insignificant. From the numerous possibilities only two correlations of any interest emerged and both of these were for the selected "after" group which will be remembered as being characterized by selective adoptive placement based on the infant examination. For the selected "after" group the psychologist's rating of Sensitivity to Failure (made at the time of our follow-up study) was correlated to the order of .37 with the Personal-Social developmental quotient of the infant examination. The corresponding correlation for the selected "before" group was .19. In view of our a priori considerations and the results of our analyses with the Binet IQ, we are not surprised to find that this correlation of .37 for the selected "after" group was reduced to the order of .24 when differences in Ambition–No Ambition and Age–Duration of Marriage were eliminated from this correlation by the partial correlation device.

A similar pattern of relationships was observed for a special composite score which we prepared, a score comprising the child's approach to the pediatric examiner, his reaction to undressing, interest, affect, and activity at the time of the physical examination during the follow-up study. This composite score was correlated with the Personal-Social developmental quotient on the infant examination to the order of .31 for the selected "after" group. The correlation for the selected "before" group was −.10. When differences in the Ambition–No

Ambition composite score were eliminated from the correlation of .31 by the partial correlation device, the correlation was reduced to .27. It is probable that this correlation could have been further reduced by the use of other variables.

It should be noted that no promising correlations emerged in the selected "after" subsample for our *older* children. We examined the data for the older children and learned that the correlations between aspects of the infant examination and aspects of the environment were not as high as the respective correlations of the younger sample. Although the difference between the correlations for the two age groups would not have reached the criterion for statistical significance, the difference in magnitude may be sufficient to account for the fact that the correlations between the infant examination and the criteria did not emerge. Some of our subsequent analyses suggest an explanation. In general, the standards for selecting the adoptive homes were higher prior to 1945 when the older sample was being placed than subsequent to 1945 when the younger sample was being placed. The difference, particularly with respect to such variables as ambition (which, incidentally, was not a factor in the selective placement of the older children) were not differences in range but differences in level. It seems possible, therefore, that some of the conspicuously less favored homes which might have had a detracting effect on the child's development were not represented in our older sample to an appreciable degree. This is not a necessary interpretation, however. It is possible that the *school child* is not dependent in his development on the home in the same manner as the *preschool child*. In this connection, however, it may be remarked that both the younger and the older sample provide evidence of relationship between aspects of the environment and aspects of development. More will be said of this in our report on the environmental relationships.

CONCLUSION

In conclusion, our exploration of the relationships between the numerous facets of the infant examination and the numerous criteria for development which we have employed has yielded a lengthy and monotonous series of correlations which were not only unreliable but which were too small to be of any practical interest had they been based on huge samples capable of making practically insignificant correlation coefficients statistically significant. The correlations between aspects of the infant examination and our criteria which were statistically significant and large enough to suggest some practical value were so infrequent as possibly to have occurred as a result of chance alone. Moreover, these more interesting correlations occurred in a subsample where selective placement on the basis of the infant examination was not only anticipated but was consistent with correlations between the infant examination and certain conspicuous characteristics of the adoptive parents which could have been known to the placement agencies. Partial correlation analyses sustained the selective placement hypothesis for this questionable group.

Aside from the few correlations produced by the subsample where selective placement based on the infant examination was conspicuous, the other interesting correlations were for strength of grip and weight quotient. These correlations were not found for all subsamples, and since the physical examination was not conducted for our older group, they could not be confirmed. They are far from conclusive and only suggestive, but their implications should be explored by analyses which emphasize physical aspects of development.

Our investigations have sought evidence of predictive validity both with respect to facets of the infant examination which other investigators have not examined and with respect to criteria which other investigators have not employed. (Some of our criteria, such as ratings and cluster scores, are not as well tried or as reliable as our mental ability, scholastic achievement, or physical development criteria. Nevertheless, our cluster scores, and by inferences from numerous significant correlations, our ratings, too, offer a modicum of reliability and should be sufficient to indicate the presence of predictive validity despite the possibility that they would not yield suitably exact indications of *degree* of validity.) Despite the broad basis of our study, we, like the others, have no basis for challenging the hypothesis that the infant examination is without predictive validity. As we have frequently emphasized, however, such a result does not *prove* that the infant examination has no predictive validity, and one must remember that it is never possible to prove a hypothesis of no relationship.

APPENDIX A

TEST REACTION SCALE

Mother-Child Relationship as Observed by Examiner

1. Child was not at all hesitant about leaving mother, scarcely said goodbye and went with examiner readily. He showed no concern for mother during examination and was extremely casual about seeing his mother again at close of examination.
2. Child was somewhat hesitant at leaving mother, said goodbye and looked to see where she would wait, and showed pleasure at seeing her again at close of examination.
3. Child was extremely reluctant to leave mother, held on to her hand, may have cried, wanted her to come with him, needed to be reassured that she would wait for him. He asked about her several times during the examination. He seemed very much relieved to see her again.

Attention

1. Completely absorbed by task with no interference from distracting stimuli.
2. Almost always absorbed by task but a little interference from distracting stimuli.
3. Pays attention to task. Efficiency is not impaired by attention to outside stimuli.
4. Child is easily distracted by extraneous stimuli or by his own ideas, but returns readily to the task.
5. It is difficult to get and maintain the attention of the child.

Activity

1. Child can't sit still for any length of time. Often stands up, explores the room, shifts position, climbs around the chair.
2. Child generally sits still for the interview but may occasionally become restless, shift position, slide down in the chair.
3. Child sits still during most of the interview unless told to stretch or stand.

Sociability

1. Child adjusted to the examination with immediate friendliness. He smiles, laughs, shares experiences with the examiner shortly after meeting.
2. It requires a while for him "to warm up" to the examiner.
3. The child maintains an attitude of reserve for a considerable length of time. It is only toward the close of the examination that any friendliness is manifested.
4. The child at no time becomes friendly, laughs, smiles or talks spontaneously or freely with the examiner.

Belligerence

1. Child is markedly belligerent. He attempts to strike or insult the examiner or to mar equipment.
2. Child is surly and may deliberately refuse to cooperate.
3. Child displays no evidence of hostile reactions during examination.

Maintenance

1. After the child feels at ease with the examiner, favorable cooperation and motivation are sustained for the duration of the contact.
2. After the child feels at ease with the examiner favorable cooperation and motivation are generally sustained but the child may have moments of apprehensions, distrust or belligerence.
3. After the child feels at ease with the examiner favorable cooperation and motivation deteriorate so that the child cries, refuses to continue, becomes excessively fatigued.

Reaction to Failure

1. Child reacts to a feeling of failure with marked increase in effort and persistence.
2. Child reacts to a feeling of failure with a slight increase in effort and persistence.
3. Child reacts to a feeling of failure with slight decrease in effort and persistence.
4. Child reacts to a feeling of failure with a marked decrease in effort and persistence.

Sensitivity to Failure

1. The child is oblivious to his failures and indifferent to any suggestion of error.
2. The child shows some recognition of his errors and slight concern about errors.
3. The child is markedly sensitive to failure and shows considerable concern about errors.

Male Versus Female Examiner

1. Child exhibits a greater degree of enthusiasm, friendliness and confidence with a male examiner than with a female examiner when other factors remain the same.
2. Child exhibits the same degree of enthusiasm, friendliness and confidence with a male examiner as with a female examiner when other factors remain the same.
3. Child exhibits a lesser degree of enthusiasm, friendliness and confidence with a male examiner than with a female examiner when other factors remain the same.

Modesty

1. The child removes his clothes readily.
2. The child removes his clothes only with encouragement and persuasion.
3. The child is extremely reluctant to remove his clothes.

Thumb Sucking

1. During the time the child is observed at the clinic he sucks his thumb whenever otherwise not engaged.
2. During the time the child is observed at the clinic he sucks his thumb intermittently.
3. During the time the child is at the clinic he is observed to suck his thumb.
4. During the time the child is observed at the clinic there is no thumb sucking.

Fine Motor Ability

1. The child exhibits continuous and precise controls so that fine motor activities are rapid and accurate.
2. The child exhibits some control over fine motor activities but movements are occasionally inaccurate or somewhat slow.
3. The child exhibits little control over fine motor activities so that movements are inaccurate or clumsy.

Speech Defect

1. The child has a sufficiently severe speech defect that interferes with communication.
2. The child has a mild speech defect which, however, does not interfere with communication.
3. There is no evidence of speech defect.

APPENDIX B

A.C.R.P. EXAMINATION FORM

Name ______________________ Sex ______ Birth Date ______ Date of Ex. ______

Height ______________ Weight ______________

Pertinent History:

Head: Circumference ______________

Shape: Dolichocephalic ______________

Brachycephalic ______________

Other ______________

Hair: Color ______________ Consistency ______________

Eyes: Color ______ E.O.M. ______ L.&A. ______ Fundus ______

Vision: R L

Other

Ears: External ______________ T.M. ______________

Hearing: R L

W.V. (15″)

Nose:

Mouth: Teeth: Number ______ Caries ______ Fillings ______

Orthodontic ______

Tongue:

Mucosa:

Throat: Tonsils: Present ______________ Absent ______________

Size ______________

Neck: Adenopathy:

Thyroid:

Chest: Shape:

Lungs:

Heart & Circulation:

Rate ______________

Rhythm ______________

Murmurs ______________

Size ______________

Femoral Pulse ______________

Abdomen: Shape:

Liver: Kidney: Spleen:

Tenderness:

Masses:

Genitalia: Circumcision: Other:
Testes Descended:

Skin:

Ext. Back Spine ______________ Deformity ______________

Joints ______________ Limitation of Motion ______________

Neurological:

Cranial Nerves ______________
Motor Nerves ______________
Sensory Nerves ______________

Reflexes: 1. Hyper 2. Normal 3. Hypo 4. Absent

	R	L
T		
B		
BR		
Abdomen:		
Upper		
Lower		
K		
A		
P		
Crem.		

Gross Motor:

Agile
Normal
Clumsy

Handedness: R ________ L ________

Strength: Dyn. ______________
Jump ______________
Pull ______________

Personality

Approach to Examiner
________ 1. Spontaneous
________ 2. With suggestion
________ 3. With urging
________ 4. Resistive

Speech
________ Spontaneous
________ Hesitant
________ Lisping
________ Stammer
________ Infantile
________ Other

Reaction to Undressing
________ Unassisted
________ Partly Assisted
________ Assisted
________ Prompt without show of concern

__________ Cooperative but not enthusiastic
__________ Great resistance

General Description:
Interest and Curiosity:
Affect:
Activity:

General Impression:

REFERENCES

1. ANDERSON, L. D. The predictive value of infancy tests in relation to intelligence at five years. *Child Develpm.*, 1939, **10**, 203-212.
2. ARTHUR, GRACE. *A point scale of the performance tests,* (2nd ed.) New York: Commonwealth Fund, 1943.
3. BAYLEY, NANCY. Mental growth during the first three years. *Genet. Psychol. Monogr.*, 1933, **14**, 1-92.
4. BRENNER, RUTH F., *et al. A follow-up study of adoptive families.* New York: Child Adoption Research Committee, Inc., 1951.
5. CATTELL, PSYCHE. *The measurement of intelligence of infants and young children.* Lancaster: Science Press, 1940.
6. GESELL, A. The stability of mental growth careers. *Yearb. Nat. Soc. Stud. Educ., Part II,* 1940, **23**, 149-160.
7. GESELL, A., & AMATRUDA, CATHERINE S. *Developmental diagnosis.* New York: Hoeber, 1941.
8. GESELL, A., THOMPSON, HELEN & AMATRUDA, CATHERINE S. *The psychology of early growth.* New York, Macmillan, 1938.
9. GRAY, W. S., *Gray's Oral Reading Test.* Bloomington, Ill.: Public School Publishing Company, 1915.
10. ILG, FRANCES L., & AMES, LOUISE B. Child behavior. *New Haven Register,* Feb. 3-5, 1953.
11. KELLEY, T. L., RUCH, G. M., & TERMAN, L. M. Stanford Achievement Test—Form D. Chicago: World Book Company, 1940.
12. TERMAN, L. M., & MERRILL, MAUD A. *Measuring intelligence.* Boston: Houghton Mifflin, 1937.

(Accepted for early publication October 18, 1955)

No. 410
1956

Vol. 70
No. 3

A Study of Adoptive Children:

III. Relationships Between Some Aspects of Development and Some Aspects of Environment for Adoptive Children

By

J. R. Wittenborn and Others

Rutgers University

Vol. 70, No. 3 Whole No. 410, 1956

Psychological Monographs: General and Applied

A Study of Adoptive Children: III. Relationships Between Some Aspects of Development and Some Aspects of Environment for Adoptive Children[1]

J. R. Wittenborn
Rutgers University

With the collaboration of:
Myrtle A. Astrachan, Ph.D., Psychologist
Marjorie W. DeGooyer, M.S., Psychologist
W. Wallace Grant, M.D., Pediatrician
Irma Z. Janoff, Ph.D., Psychologist
Robert B. Kugel, M.D., Pediatrician
Barbara J. Myers, M.S., Psychologist
Annaliese Riess, Ph.D., Psychologist
Ellery C. Russell, M.S., Social Worker

INTRODUCTION

THE FOLLOWING pages present the third phase of a rather lengthy investigation of adoptive children. The first phase, *"Interviews as a source of scores for children and their homes"* (17),[2] described our efforts to provide some measuring devices which were appropriate to our interests and in keeping with limitations imposed by the nature of our total investigation. The second phase, *"The predictive validity of the Yale Developmental Examination of Infant Behavior"* (18) has been described in terms of our efforts to find a useful degree of predictive significance for the infant examination; our efforts in this direction were unsuccessful.

The present phase suggests respects in which characteristics of children's environment may be correlated with characteristics of their development. Specifically, the present report is concerned with a general question which was phrased in the following manner at the time our study was planned and support was sought from the Russell Sage Foundation:

"What characteristics of adoptive homes (placements) may be shown to be correlated with characteristics of adoptive children?"

---

[1] The staff of the project wishes to express its appreciation to Dr. Donald Young, Director of the Russell Sage Foundation, and to Dr. Carl Hovland and Dr. Milton Senn of Yale University. The nature and the quality of this effort have been greatly influenced by such friends and advisers as Dr. Joseph Baldwin, Dr. Nancy Bayley, Dr. Harold Jones, Dr. Neal Miller, Dr. Robert Sears, and Dr. Helen Thompson. We are indebted also to Dr. Leonard Cottrell, Dr. Liselotte Fischer, Dr. Else Frenkel-Brunswik, Dr. Marjorie Honzik, Dr. Mary Cover Jones, Dr. Frederick Mosteller, and Dr. A. J. J. Van Woerkom. We are particularly grateful to the adoptive families who participated in this investigation, the child placement agencies who cooperated with us and guided our efforts, and to Adelaide Piscitelli, David Macklin, David Jenkins, and Peter Field who helped in the analysis of the data and the preparation of the manuscript.

[2] References for the present report are listed on page 231. The pagination of the present report is continuous with the second report of this series, entitled "A study of adoptive children: II. The predictive validity of the Yale Developmental Examination of Infant Behavior" (*Psychol. Monogr.,* 1956, Vol 70, No. 1 [*Whole No. 409*]).

It should be noted that this question requires that we consider a wide range of measures of the child's development and of the child's environment. We obviously are not committed to an "explanation" of any relationships that we find nor are we committed to a precise determination of the ultimate magnitude or degree of any of the relationships suggested by our data.

Despite the exploratory nature of our investigation, its results have implications for three related topics of current interest: the placement practices of child welfare agencies, the child-training practices of parents, and theories and beliefs concerning the development of the personality. Because of the implications of the study, we sought to conduct our central analyses with respect to two samples. One sample comprised 114 adoptive children who were in their fifth or sixth year of life but who were not yet involved in the work of the first grade. Our second sample comprised 81 adoptive children who were past the work of the first grade but were not yet preadolescent in their development; seventy per cent of these children were in their eighth or ninth year of life. It was hoped that these two samples would provide mutual confirmation for some of the readily confirmable trends and suggest qualifications for trends that most require qualification. We have been greatly concerned with the manner in which observed trends or relationships may be interpreted. Since our analyses describe relationships between fairly definable characteristics of the child's environment and fairly definable facets of his development, there is danger that one may overgeneralize and prematurely or incorrectly ascribe a literal, casual significance to our specific environmental variables.

Although the relationships to which we draw the reader's attention meet conventional criteria for statistical significance, we do not know how much practical significance to ascribe to them. Almost without exception they are plausible and consistent with possible implications of current theories of personality development. Nevertheless, the relatively crude, untried nature of some of our variables, particularly our measures for the environment, does not permit us to make confident statements concerning the degree or the importance of the relationships suggested by our data. Before we can be confident of the practical importance of these relationships, confirmatory studies should be undertaken. Such studies should draw upon samples which differ in certain respects from ours and should make use of hoped-for refinements in our measuring devices.

The purpose of our investigation requires that we examine the relationships between numerous aspects of the child's development and numerous aspects of his environment. The breadth of our interest in these two classes of variables is indicated in our descriptions of the first and second phases of our total investigation, *"Interviews as a source of scores for children and their homes"* and *"The predictive validity of the Yale developmental examination of infant behavior."* For the purposes of the present report we are interested in describing relationships for the variables which are most important and have the greatest general interest. We have considered as unimportant those variables which have fewer statistically significant relationships than could be expected on the basis of chance alone and those whose evidence of relationship is with respect to other variables of obscure or restricted importance. Variables that are uniquely relevant to the adoptive situation and may have little general interest are also eliminated from the present report.

The Data

Let us first review the measures for the environment which are employed in the present report. One important aspect of the child's environment concerns the child-rearing practices of the home. As described in our report, *"Interviews as a source of scores for children and their homes,"* we conducted lengthy semistructured interviews with the children's mothers, and during the course of these interviews gathered a standard body of specific information concerning child-rearing practices. Specific questions, however, were avoided. On the basis of both statistical and psychological criteria, the various bits of information were organized into clusters which enabled us to score the interview material with respect to such phases or qualities in child rearing as eagerness on the part of the mother, general lack of sympathy toward the child, severity of toilet training, and others included in the outline (Table 1) provided below.

In addition, the interview with the mother was extended to include the values, the organization, and the general atmosphere of the home. Information relevant to these aspects of the home was gathered by our social worker also, who visited the homes as our first direct contact with the adoptive parents. This information was organized into clusters and provided the basis for cluster scores descriptive of the home per se; we refer to these cluster scores as Conditions of Child Rearing. They are included in our outline (Table 1) below.

In addition, we were able to employ measures based on simple factual information. One of these, the *Educational-Occupational Composite Score,* is defined as follows: If the adoptive father's occupation was business or professional or in the white-collar group, a weight of plus one was arbitrarily assigned. If, however, his occupation was unskilled, semiskilled, or skilled, the weight was minus one. A weight of zero was assigned in all situations where we did not know the adoptive father's occupation. The father's education was combined by allowing plus one if he had attended college, minus one if he had not attended college, and zero when we did not know what his educational background was. The mother's education was weighted in a similar manner. These three sets of weights were combined algebraically to form a general composite score representing the educational-occupational status of the home. Another score, which we consider to be an indication of the strength and stability of the home, is called the *Age–Duration of Marriage Composite Score* and is defined in the following manner: If the adoptive father was in his thirties, a weight of plus one was arbitrarily given; if he was under thirty or forty or over (5), a weight of minus one was given; if his age was unknown, a weight of zero was given. The same weighting system was used for the age of the adoptive mother. In addition, if the couple had been married seven or more years, a weight of plus one was given; if they had been married less than seven years, a weight of minus one was given; and if the duration of their marriage was unknown to us, a weight of zero was given. These three weights were combined algebraically to give us a composite measure of the age of parents and duration of their marriage.

Because of its general interest and apparent importance, we have included the

TABLE 1

A Summary of Environmental Variables

	Younger Sample		Older Sample	
Child-Rearing Practices Clusters (based on interview with the mother)	Eagerness	.57*	Overcontrolling	.32*
	Unsympathetic	.44	Unsympathetic	.33
	Severe Toilet Training	.44	Severe Toilet Training	.45
	Punishment for Aggression	.78	Punishment for Aggression	.82
Conditions of Child-Rearing Clusters (based on interview with the mother and on social worker's visit to the Home)	Ambition	.78	Ambition	.77
	Harmonious Family Relationships	.58	Harmonious Family Relationships	.60
	No Ambition	.72	No Ambition	.74
	Rejection of Child	.76	Rejection of Child	.68
	Parental Incompatibility	.73	Parental Incompatibility	.52
	Overprotectiveness	.77		
Compositive Scores (based on data descriptive of adoptive parents)	Educational-Occupational Composite		Educational-Occupational Composite	
	Age-Duration of Marriage Composite		Age-Duration of Marriage Composite	
Adoptive Placement	Age at adoptive placement		Age at adopitve placement	

* Split-half reliabilities for cluster scores; Spearman-Brown formula.

age of adoptive placement of the child as one of the environmental variables. These variables are all summarized in the outline in Table 1, and split-half reliabilities are provided for the cluster scores. It is suggested that the reader refer to our report, *"Interviews as a source of scores for children and their homes"* (17), which provides a detailed description of our procedure for building cluster scores.

Our efforts to secure appropriate measures for the children's development are described in detail in our preceding reports (17, 18). We employed the conventional tests for mental, scholastic, and physical development. In addition, we made numerous behavior ratings and also constructed numerous cluster scores which, it was hoped, would give us some standard estimates of personal and social aspects of the children's development. One set of cluster scores was based on an interview with the children. This interview (the Social Reaction Interview) was of a simple question-and-answer type and concerned the children's evaluation of their own behavior. Another set of cluster scores was based on the mother's description of the child and emerged from the Child-Rearing Practices interview.

The most important and the most interesting of our measures of the children's development are summarized in the outline in Table 2. Some of these, such as the intelligence tests and the scholastic achievement tests, are tried and true and known to be both reliable and valid. Many of the aspects of the child's physical development secured as a part of a pediatric-neurological examination may also be claimed to be reliable and valid, but they offered so few evidences of relationship with our environmental variables that (with the exception of one variable, a Weight Quotient found by dividing a child's weight by his age) they are excluded from our summary. Our various measures for the children's personal-social development are new and are less reliable than the tests of intelligence. Nevertheless, they have sufficient reliability to indicate the presence of relation-

TABLE 2

A SUMMARY OF THE DEVELOPMENTAL VARIABLES

Criterion	Younger Sample		Older Sample	
I. Intellectual development	Stanford-Binet IQ Arthur IQ		Stanford-Binet IQ Arthur IQ	
II. Scholastic achievement	(none relevant)		Reading Spelling Arithmetic	
III. Motor development	(no promising relationships)		(no promising relationships)	
IV. General physical development (based on neurological-pediatric examination)	Weight quotient		(no examination given)	
V. Personal-social development	Dependence on adults	.88*	Cooperation with authority	.41*
A. Child's reaction to himself—the Social Reaction Interview	Socialized compliance	.70		
	Takes adult role	.69		
	Weakness-avoidance	.57		
B. Mother's reaction to the child—Child-Rearing Practices Interview	Compulsive	.52	Aggressive	.64
	Aggressive	.70	Anxious	.56
	Phobic	.57	Dependent	.60
	Dependent	.71	Anxious-aggressive	.80
			Spoiled child	.78
			Controlled	.78
C. Ratings made by examiners	Sensitivity to failure		Sensitivity to failure	

* Split-half reliabilities for cluster scores; Spearman-Brown formula.

ships, where definite relationships exist. Although the measures of personal-social development have the obvious handicap of uncertain reliability and of newness, they are possibly as desirable as any other currently available measures; rating procedures for relevant phases of the child's development have been provided by Baldwin (2), but his procedures were infeasible in our study of adoptive families.

Particular consideration should be given to the adoptive status of our children and their families. This circumstance has certain advantages. It makes our results directly relevant to certain child-placement practices and of immediate interest to child-placement workers. It also suggests that some of our relationships between aspects of the child's development and aspects of his environment may be independent of heredity as a confounding third variable. Our data may, therefore, add convincingly to the considerable body of evidence relevant to the importance of environmental factors in the child's development, particularly those aspects of development which are often called personal and social.

Readers will seek practical interpretations for our results. Accordingly, we must explicitly indicate certain possible sources of confounding which could lead to alternative interpretations. We have considered three major sources of possible confounding in the relationships between characteristics of the child's environment and characteristics of his development.

1. Our subjects were adoptive children living in adoptive homes. It is possible that knowledge of the true parents was a factor in the selective placement of the adoptive children; we have evidence which favors this possibility. If this knowledge of the true parents has an hereditary significance, one would expect respective hereditary attributes of the children to be correlated with any characteristics of the adoptive homes which were selected on the basis of knowledge of hereditary attributes of the true parents. For some of our trends relating characteristics of the child with character-

istics of his environment, this alternative explanation is plausible and an hereditary hypothesis is a reasonable alternative to an environmental hypothesis.[3] Other relationships involve characteristics of the environment and characteristics of the child for which no hereditary hypothesis is generally recognized; for these relations an *ad hoc* hereditary hypothesis would be particularly implausible because those who mediated the adoptions could not have had the relevant information concerning either the true parents or the subsequent practices of the adoptive parents.

2. Despite our failure to find evidence for the predictive validity of the infant examination (18), it is possible that some readers believe that the infant examination had for our criteria a predictive validity which we failed to discover. Since the infants were placed selectively as a result of the infant examination, some of the correlations between aspects of the adoptive home and aspects of the child's development could emerge as a result of the fact that both the environment (as a result of selective placement) and criteria of development (as a result of putative validity of the baby test) are correlated with the infant examination. We have conducted analyses which are relevant to this possible source of confounding. Specifically, the relationships between aspects of the environment and aspects of development which are revealed in our total group were studied in a subgroup where the infants were placed adoptively *before* the infant examination. In this subgroup, where no selective placement could have resulted from the infant examination, we still find the general pattern of relationships between aspects of the child's environment and aspects of his development.

3. As a third source of confounding, one might wonder whether some of the relationships between descriptions of the child and descriptions of the home provided by our interviews are not correlated because of their common source, either because of bias on the part of the mother or because of bias on the part of the interviewer. We have three reasons for doubting that this is an important source of confounding:

(*a*) Neither the mother nor the interviewer knew how the various bits of information were to be combined to form cluster scores and we greatly doubt if any a priori considerations would have been sufficient to yield the particular pattern of results we have.

(*b*) We find that when the descriptions of the child emerging from the psychologist's interview with the mother are correlated with the same kind of environmental information from two *different* sources, viz., the social worker's visit to the home and the psychologist's interview with the mother, the pattern of relationship is uniform and essentially unchanged. (Obviously a relationship between the child and the home is not dependent upon a common source of information.)

(*c*) The general sense of the relationships between characteristics of the home as they are known through the interview with the mother and characteristics of the child as they are known from the interview with the mother is not challenged wherever other sources of relevant information concerning the child and major objective facts concerning the home are available.

---

[3] A sample of adoptive children placed at random in adoptive homes could provide a basis for a conclusive test for this hypothesis.

THE CORRELATIONS

Table 3 provides a summary of the correlations between the most promising criteria of the child's development and the most interesting measures of the child's environment; these data are for the younger sample. The results for the older sample are presented in Table 4. Scrutiny of these two tables reveals many similarities between them with respect to their over-all implications and provides some assurance that the trends indicated are not peculiar to one sample alone, and are not the capricious result of sampling fluctuations. As we shall note in our subsequent discussions, the implications of some of these correlations are in good agreement with the implications of relationships described by other investigators.

The absolute magnitudes of the relationships indicated by our data are relatively small. This, however, may be in part an expression of the crude nature of the variables which we were obliged to use in our exploration. The hypothesis that the intrinsic relationships suggested by our data may in truth be strong is worthy of the most serious consideration and should be explored by means of refinements in the devices we have used.

As already indicated, the reader may be concerned with the possibility of certain sources of confounding. Therefore we must point out the situations where we consider a possibility of confounding to exist and, wherever possible, we must eliminate confounding as an alternative explanation for the trends we describe.

Let us consider first those relationships where a confounding with heredity may be suspected. Heredity may be suspected as a common third variable in our relationships between criteria of development and measures of the environment when the aspect of a child's development is commonly suspected to be an hereditary expression of a comparable attribute in the parent. This is particularly important when both the environmental facets (which we have found to be correlated with this putative hereditary aspect of development) and the putative hereditary attribute of the parent could have been known by those who were involved in the adoptive placement of the infant; stated somewhat differently, an hereditary explanation for our trends may be considered in those cases where an attribute of the true parent is known by those responsible for the child's placement and where the relevant features of the adoptive family could have been known so that selective placement based on knowledge of the true parents could have occurred.

Among our measurements of development those aspects which possibly are most commonly considered to have an hereditary basis could be the Binet IQ, the Arthur IQ, and the Weight Quotient. The correlated aspects of the child's environment which might have been known by the placement workers (or could have in some way been factors in selective placement in a manner relevant to the brightness or the size of the true parents) could be expressed by the cluster scores Ambition and No Ambition and the composite scores Educational-Occupational Status and possibly Age–Duration of Marriage. The ambiguous status and the uncertain meaning of the relationships between the mental aspects of the child's development and such features of his environment are a familiar problem to psychologists. The correlations obviously describe relationships between the environment and the child's development and one may hypothesize that the relationship is an expression of

TABLE 3

A Summary of the Reliable Relationships Between Criteria* for the Child's Development and Some Measures of His Environment

(Younger sample; $N = 114$**)

	Child-Rearing Practices Clusters				Conditions of Child-Rearing Clusters						Composite Scores		
Criteria of Child's Development	Eagerness	Unsympathetic	Severe Toilet Training	Punishment for Aggression	Ambition	Harmonious Family Relationships	No Ambition	Rejection of Child	Parental Incompatibility	Overprotectiveness	Education-Occupation Composite	Age-Duration of Marriage Composite	Age at Adoptive Placement
I. Intellectual development													
A. Stanford-Binet IQ	.20	—	—	—	.27	—	−.26	—	—	—	.31	.39	−.35
B. Arthur IQ	—	−.32	—	−.22	—	—	—	−.31	—	—	.19	.27	−.27
II. Scholastic achievement (older sample only)													
III. Motor development (no promising relationships)													
IV. General physical development													
Weight Quotient (no other promising relationships)	—	—	—	—	—	—	−.20	—	−.28	—	—	.19	—
V. Personal-social development													
A. Child's reaction to himself (Social Reaction Interview with the child)													
Dependence on adults	—	—	—	—	—	—	—	.30	—	—	—	—	—
Socialized compliance	—	—	—	—	—	—	—	—	—	—	.26	.50	—
Takes adult role	—	—	—	—	—	—	—	—	—	—	—	.32	−.50
Weakness-avoidance	−.19	—	—	—	—	—	—	—	—	—	—	.27	—
B. Mother's reaction to the child (Child-Rearing Practices Interview with the mother)													
Compulsive	—	.37	.31	—	—	−.20	—	.24	.27	—	—	—	—
Aggressive	—	.42	.36	.31	—	—	—	.24	—	—	—	—	−.24
Phobic	—	.33	.22	.23	—	—	—	.23	.24	—	—	—	—
Dependent	—	—	—	—	—	−.22	.21	—	—	.26	—	—	—
C. Ratings made by examiners (Test Reaction Scale filled in by psychologist)													
Sensitivity to Failure (no other promising relationships)	—	—	—	—	.21	—	—	—	—	—	.28	—	—

---

* Criteria which showed no more evidence of relationships than could be ascribed to chance were eliminated from consideration.

** r significant at the .05 level $\geq .19$; r significant at the .01 level $\geq .25$.

TABLE 4

A SUMMARY OF THE RELIABLE RELATIONSHIPS BETWEEN CRITERIA* FOR THE CHILD'S DEVELOPMENT AND SOME MEASURES OF HIS ENVIRONMENT (Older sample; $N=81$**)

	Child-Rearing Practices Clusters				Conditions of Child-Rearing Clusters					Composite Scores		
Criteria of Child's Development	Over-Controlling	Unsympathetic	Severe Toilet Training	Punishment for Aggression	Ambition	Harmonious Family Relationships	No Ambition	Rejection of Child	Parental Incompatibility	Education-Occupation Composite	Age-Duration of Marriage Composite	Age at Adoptive Placement
I. Intellectual development												
A. Stanford-Binet IQ	—	—	.29	.26	.25	—	—	—	—	.30	—	−.26
B. Arthur IQ	.30	—	.29	—	—	—	—	—	—	.39	—	−.37
II. Scholastic achievement												
Reading	—	.22	—	—	—	—	—	—	—	—	—	−.25
Spelling	—	—	—	—	—	—	—	—	—	—	—	−.58
Arithmetic	—	.22	—	—	—	—	—	—	—	—	—	—
III. Motor development (no promising relationships)												
IV. General physical development (younger sample only)												
V. Personal-social development												
A. Child's reaction to himself (Social Reaction Interview with the child)												
Cooperation with Authority (no other promising relationships)	—	—	—	−.27	—	—	—	—	—	—	—	—
B. Mother's reaction to the child (Child-Rearing Practices Interview with the mother)												
Aggressive	—	.25	.30	.26	—	—	−.28	.31	—	—	—	—
Anxious	—	.35	.42	—	—	—	—	.41	—	—	—	—
Dependent	.24	—	—	—	—	—	—	—	−.34	—	—	—
Anxious-aggressive	—	.27	.36	.33	—	—	−.27	.36	.36	—	—	—
Spoiled child	—	—	—	.42	—	—	—	—	—	—	—	—
Controlled	—	—	−.31	—	—	—	—	−.29	—	—	—	—
C. Ratings made by examiners (Test Reaction Scale filled in by psychologist)												
Sensitivity to Failure (no other promising relationships)	—	—	—	—	.24	—	—	—	—	—	—	—

---

* Criteria which showed no more evidence of relationships than could be ascribed to chance were eliminated from consideration.

** r significant at the .05 level $\geq .22$; r significant at the .01 level $\geq .29$.

differences in environment per se. One may also hypothesize that the relationships are an expression of the fact that the differences in the environment are organized, i.e., selected on the basis of differences in the true parents, and that the relationship between the environment and aspects of development is but an indirect expression of an intrinsic relationship of an hereditary nature between the true parents and the adoptive children. In keeping with this hypothesis, we have found significant relationships between the educational-occupational composite score for the true parents and the educational-occupational score for the adoptive parents (it would appear that some selective placement based on knowledge of the true parents had occurred). In addition, we have found significant relationships between the educational-occupational scores of the true parents and the Binet IQ's of the children.

This pattern of relationships is equally consistent with either an environmental hypothesis or an hereditary hypothesis. Does the correlation between the true parents' educational-occupational status and the child's Binet IQ occur merely as an incident to the fact that knowledge of the true parents had something to do with the adoptive placement (i.e., an environmental pattern which may have influenced Binet IQ) of the children? Does the correlation between the environmental factors and the Binet IQ of the children occur merely as an accident of the fact that selective adoptive placement was based in part on such knowledge of the true parents as their educational-occupational status (which in the opinion of some readers may express the kind of hereditary difference which may also find expression in Binet IQ)?

Wherever obvious possible sources of confounding may be discarded, the relationships are much more provocative both from the practical and from the theoretical point of view than where the possibility of confounding is obvious but not subject to evaluation. It will be noted in Tables 3 and 4 that we have employed our *total* younger sample and our *total* older sample in showing the relationships between the most promising criteria of development and the most promising measures for the environment. The reader may recall from our study of the validity of the infant examination (18) that in one of the subgroups of children, where the infant examination seemed to be most satisfactory to the examiner, there was indication that the infants were placed in their adoptive homes selectively so that correlations between the infant examination and certain aspects of the adoptive homes emerged. The aspects of the adoptive homes which seemed to have been selected partly on the basis of the infant examination included the Ambition cluster score, the No Ambition cluster score, the Educational-Occupational composite score for the adoptive parents, and the Age–Duration of Marriage composite score for the adoptive parents. In Tables 3 and 4 it may be noted that those aspects of the environment are correlated with the Binet IQ and that the two composite scores also show some relationship with the Arthur IQ. If (despite our failure to find any indication that the infant examination is predictive of either the Binet or the Arthur) the reader continues to believe that some important relationship nevertheless exists, he might suspect that the relationships between the Binet and those aspects of the environment are but artifacts. Specifically, he might suspect that the relationship between the Binet and these as-

TABLE 5

CORRELATIONS BETWEEN MENTAL ABILITY AND CERTAIN ASPECTS OF THE ENVIRONMENT WHICH WERE SELECTED ON THE BASIS OF THE INFANT EXAMINATION; A COMPARISON BETWEEN THE TOTAL YOUNGER GROUP OF 114 AND A SUBGROUP OF 32 PLACED ADOPTIVELY BEFORE THE INFANT EXAMINATION

Test	Group	Cluster Scores		Composite Scores	
		Ambition	No Ambition	Education-Occupation	Age-Duration of Marriage
Binet IQ	Total younger group	.27	−.26	.31	.39
	Selected-before subgroup	.36	−.36	.28	.28
Arthur IQ	Total younger group	—	—	.19	.27
	Selected-before subgroup	—	—	.25	.12

pects of the environment emerge from the fact that both the Binet and the environmental variables are correlated with the infant examination. Although we consider this possibility to be extraordinarily unlikely, readers of this persuasion will be interested in the correlations for a sample of children who were placed in their adoptive homes *before* they received an infant examination and for whom there could be no possible confounding between the infant examination and the environment of the adoptive homes. We refer to this special group where no selective placement based on the infant examination could occur as the "selected-before group." In Table 5 comparisons are provided between the correlations for this selected *before* group and the *total* group. This comparison is conducted for the younger sample only, because the "selected-before group" for the older sample was quite small. It will be seen that the correlations between the aspects of the environment and the IQ's are essentially the same for the selected-before group as they are for the total group. It is obvious that such correlations between aspects of the child's development and aspects of his environment are not dependent upon confounding with the infant examination.

It is of incidental interest that such comparisons were made for all of the correlations between aspects of environment and the various criteria of development, and the essential similarity of relationships was observed.

Since the families in which the adoption was mediated by the agencies were reached by us through the agencies, one might suspect that this possible selective influence in our follow-up could have influenced the relationships revealed by our data; i.e., one might suspect that the agencies would be more successful in helping us reach those homes where there was a "correspondence" between the home and the child than where there was not. Since our children who were placed adoptively before an infant examination were almost entirely independent placements, the data for this group provide an incidental control (see Table 5). The relationships between the environmental aspects and the developmental aspects are essentially the same for the group placed before the infant examination as they are for the total group. Other specific analyses between groups placed independently and groups placed by agencies show that in *both* groups relationships between aspects of the environment and aspects of development emerge. Although the agencies may have some selective influence in our follow-up (see the description of our sample in the preceding monograph [18]), this influence is not necessary for the trends and relationships we describe.

In the preceding section we also recognized the possibility that some of the relationships between characteristics of the child based on the mother's description and characteristics of the home conceivably could be confounded either because the mother was a common source for much of these data or because the data emerged from a given interview and could have been biased by the interviewer. Since neither the mother nor the interviewer knew how the various bits of data were to be combined in order to form cluster scores, it seems improbable that our relationships are the result of bias. Fortunately, it is possible for us to show that some of the relationships between aspects of the environment and aspects of the child's personal-social development are not necessarily dependent upon bias. For example, some of the items of the Rejective cluster score were gathered by the social worker at the time of her visit to the home and the same items were gathered also by the psychologist at the time of her interview with the mother. Despite the fact that these bits of information were gathered by different people at different times and in different contexts, the limited Rejective cluster scores from these two sources are correlated to the order of .60; both the limited Rejective cluster score based on the psychologist's data alone and the identical limited Rejective cluster score based on the social worker's data alone show the same pattern of relationships with the Compulsive, Aggressive, and Phobic cluster scores for the child. (The general sense of these correlations is indicated in Tables 3 and 4.) These particular correlations at least are not dependent upon bias or confounding at the source of the data.

The over-all pattern of the relationships which are described in Tables 3, 4, and 5 does not require us to conclude that any of the obvious sources of confounding are necessary for our relationships. We are therefore inclined to consider most seriously the possibility that these correlations indicate that the environment may play some kind of facilitating or delimiting role in the aspects of child development which we have studied. Certainly our data are in no sense a proof of this environmental hypothesis, but in view of its great practical implications for parents, child-placement workers, and others interested in the development of children, it is a hypothesis which should be most seriously considered by the skeptical and should provide some encouragement to those who employ environmental considerations when explaining, anticipating, or planning for the development of children. Nevertheless, it would be premature to reject alternative hypotheses for the observed relationships, and at least for some of our relationships alternative hypotheses are clearly possible. For example, the relationships between the measures of mental ability and our measures of the home which are indicative of socioeconomic status could very well express a confounding with hereditary influences resulting from selective placement of the infant based on knowledge of the true parents. In addition, it is possible that some confounding of the bias of the mothers or the bias of the interviewer is involved in the relationships between child-rearing practices of the home as described by the mother and the characteristics of the child as described by the mother. It should be obvious to the reader, however, that the possibility of confounding in these relationships is not clear *evidence* for confounding, and should not deter us from a serious consideration of the possible practical implications of these relationships.

It is of interest to note that there are several relationships described in Tables 3 and 4 which appear to be independent of any plausible confounding of sources of data. For example, we have no reason for considering the confounding of source of data hypothesis as an explanation for the correlations between the cluster scores based on the child's Social Reaction Interview and any of our measures for the environment. Moreover, we do not have any confounding hypothesis for the relationship between Sensitivity to Failure on the part of the child and Ambition on the part of the adoptive parent. We believe that our analyses eliminate any obvious confounding hypothesis for the relationships between the Rejective cluster score and the various aspects of the child's development. We have no confounding explanation for the suggestion that rejective, unsympathetic, and punishing influences may detract from Arthur IQ in our younger sample. With respect to the age-of-adoptive-placement variable, we see no reason for using a confounding hypothesis for the various relationships involved in aspects of the child's development. It should be unnecessary to add, however, that age of adoptive placement per se is not a very obvious explanatory concept; nevertheless it seems quite possible that environmental differences associated with age of adoptive placement may very well have the importance implied by the correlations offered in Tables 3 and 4.

The reader may be concerned with the fact that although the correlations for the younger and the older samples in general may be seen as mutually confirmatory, there are some relationships which are not confirmed. We believe that the discrepancies between the results for the two samples may be explained in terms of the differences in age of the two samples and the difference in the circumstances which accompany differences in age. There were also some slight changes in the placement practices of agencies which could contribute to differences in the two age groups.

Discussion

The correlations that have been described require further discussion. As we shall see, such a discussion is most fruitful when it is based on a brief survey of the relevant data provided by the psychological literature. In the following paragraphs we shall present a terse description of studies which are most conspicuous from the standpoint of rigor and sample size. This survey is not intended to be exhaustive, but it should serve both to confirm our empirical findings and to help us interpret them. The studies presented first have to do with mental ability, and it will be seen that the investigators, like us, are involved with the possibility that the relationship between measures of mental ability and aspects of the environment are in some manner confounded with heredity.

In the published studies describing the relationship between aspects of the child's personal-social development and aspects of his environment, the children are all living with their true parents and the question of whether or not the relationships are confounded with common heredity need not be raised because this possibility of confounding is so conspicuous.

Since it seems to us particularly improbable that heredity could be a hidden third variable to account for the relationships we have observed between various aspects of the child's environment and various aspects of his personal-social development, we are greatly interested in the possibility that the reports of others, where this confounding could have occurred, are in general confirmed by our data. If our data, where hereditary confounding seems particularly improbable, confirms the implications of these other studies, we may be increasingly confident that the environmental implications of the studies of others concerning the child's personal-social development are not necessarily an indirect expression of hereditary influences.

Any congruence between the implications of our investigation and of the investigations of others with respect to the personal-social development of children has an additional significance for our investigation. In some parts of our investigation, we may suspect that bias of the mother may have produced part of the correlation between the child-rearing practices of the mother and her description of the child. If the implications of our data are consistent with the implications of the data provided by other investigators where this particular source of confounding is improbable, we may place added confidence in that portion of our results where we might fear that the mother's bias is a possible source of our relationships.

Let us proceed with our review. In 1928 Barbara Burks published a report that revealed relationship between the intelligence of foster children and the cultural index of the foster home (6). Her study showed evidence of relationship between the child's IQ and each of the following: family income, home ownership, number of books in child's library, and the amount of home instruction given the child. Although the correlation coefficients are not sufficiently high to provide a useful basis for prediction, they are sufficient to suggest strongly that the opportunities and the stimulations provided by a child's environment may have a role in his mental development.

In 1935 Alice Leahy published a report which was similar to Burks' with respect to design, purpose, and implications (10). It should be pointed out that in both these studies there is indication that the mental ability and intellectual and other attainments of the foster parents bear a relationship with the mental ability of the foster child. It must be considered, of course, that the children may have been placed selectively in the foster homes. As a matter of fact, this is probable because, in general, persons responsible for adoptive placement attempt to match the expectations of the adoptive families with the developmental promise of the adoptive child.

The probability of such relationships as these being solely due to selective placement is reduced by several analyses described by Freeman in 1928 (7). He showed that adoptive siblings who had been reared apart were correlated in intelligence to the order of .25, whereas other investigators have found that the intelligence of siblings living together is about .50. This reduction in sibling resemblance could be taken as indication that an important part of the similarity between siblings is due to a similarity in their environments. In addition, Freeman also reports that where there are relatively great differences in the cultural rating for the foster home, the correlation between pairs of separated siblings with respect to intelligence is only .19, whereas if pairs of separated siblings went into foster homes which were similar to each other with respect to cultural rating, the correlation between such pairs of siblings was .50. The disparity in intelligence that differences in environment may produce (i.e., environmental differences as introduced in Freeman's groups of siblings placed in separate foster homes) is particularly impressive evidence for the importance of environment in the development of mental ability. This appears especially probable when it is noted that for the most part Freeman's children spent several early years living together in their original home. As a complement for this analysis, Freeman also examined the correlation between the intelligence test scores for 72 pairs of unrelated foster children who were living in the same foster home. The correlation for these unrelated foster siblings who had a common home environment was .37.

In 1931 Sims (15) reported a study which has implications very similar to those provided by Freeman's analyses. He formed 203 pairs of children who were unrelated but who were living in homes matched with respect to socioeconomic status as measured by the Sims Score Card (and matched with respect to certain other factors not directly relevant to our present interests).

The correlation betwen IQ's for these 203 pairs of unrelated children was .35, i.e., unrelated children who have similar homes tend to have similar IQ's.

In both Freeman's and Sims' studies it is possible that the children in the same homes were selected in some manner so that their "heredity" was similar, and that Freeman's correlation of .37 and Sims' correlation of .35 should be ascribed to similar environment plus similar heredity. From the practical standpoint, however, we do not know that the heredity was similar. Nevertheless, we do know of some similarity in the environments.

Thus we see that, at least with respect to measured mental ability, environment appears to be an important correlate (the actual determiner is unknown). As may be inferred from the preceding paragraphs, it has not been difficult for investigators to show evidence of relationship between mental ability and facets of the socioeconomic and educational attainments of the child's home. The possibility of showing such relationships is obviously greatly favored by the availability of reliable measures of the mental ability of children. Similarly, some of the aspects of homes where primarily mental attainments are emphasized may be specified rather objectively. In addition, relatively satisfactory measures of the socioeconomic status of homes have been available for a number of different kinds of investigations. Accordingly, it is not surprising that a reliable relationship between the mental ability of the child and some characteristics of the home could be revealed.

With respect to the features of the child's behavior which are often described under the general heading of character, temperament, personality, or adjustment, the research reports often leave much to be desired. Nevertheless, there is a large literature which offers consistent support for a general hypothesis that there are important relationships between a child's personality and his environment (particularly the environment which comprises the reaction of his parents toward him). Although the published reports corroborate this principle, their content is diversified and as a consequence there are not many which are mutually corroborative in specific respects. This is almost a necessary consequence of the fact that innumerable facets of a child's personality may be of interest to investigators and almost countless aspects of the child's environment may also be the subject of investigators' interests. For this reason the probability of studies overlapping greatly is not large and consequently the possibility of independent corroborative studies is small. Since neither the most relevant aspects of the child's personality nor the most relevant aspects of a child's environment is a matter of general agreement, devices for measurement which are well developed and generally understood have not yet emerged in this area of inquiry. As a matter of fact suitable indices of a child's personality or of his interpersonal environment are relatively rare. Many of the devices that have been employed are of uncertain reliability, have a meaning which is unique to some special application, or appear to be so heterogeneous in content as to have no specifiable implications. Despite the handicap of great diversity of interest among investigators and of our general impoverishment from the standpoint of measuring devices, studies have emerged which as a group have strongly compelling implications for the importance of the environment as a possible factor in the adjustment of the child.

Only a small and relatively new literature for this field has the merit of usefully large samples, specifiable procedures, and suitable consideration for control. Fortunately, however, the small scientific and directly relevant literature exists in the context of a huge clinical, anecdotal-speculative, or analogical-experimental literature which in its implication is in general agreement with the more scientific applicable literature. In the following paragraphs a few of the more systematic studies will be mentioned briefly. Perhaps it will be of some interest to the reader if we present the selected studies in the order of their publication.

One of the earliest studies to attract our interest serves as a transition from our foregoing emphasis on intellectual factors to an emphasis on the relationship between interpersonal aspects of a child's environment and the child's personal development. In 1933 David Levy described an investigation wherein eight conspicuously overprotected children were compared with other children on the basis of scholastic achievement (12). In brief he found that the overprotected children were up to standard or possibly accelerated with respect to reading, but were substantially retarded in arithmetic. This difference is simply and plausibly explained in terms of features of the home environment of the overprotected children.

In some cases, data which have been secured in a remote context provide analyses which are of direct interest to our problem. For example, in 1934 Goodwin Watson described an analysis of the responses of graduate students to a questionnaire that had to do with the strictness of the student's home training and with some of the student's personal characteristics (16). It was revealed that students who described strict practices by their parents also reported numerous characteristic tendencies for themselves. Among these were tendencies to be ashamed of their parents, to hold grudges, and to feel that people had been unfair to them. They described in themselves a tendency to be bashful, picky about food, and curious about sexual matters. They also tended to worry. The foregoing tendencies were less characteristic of the graduate students whose responses indicated that the home training was lax. In general this report suggests that strict home training may be associated with the development of aggressive reactions and attitudes, by the development of anxiety characteristics, and by compulsive and dependent manifestations.

In 1936 Newell examined the record of 75 clinic cases of maternal rejection and compared these cases with 82 nonclinic cases (13). Although his methods of description are more verbal than quantitative, he concludes that if parental rejection is expressed by hostile handling of children, the children tend to be aggressive; expression of parental rejection by consistently protective parental handling was associated with submissive behavior on the part of the children. Newell also points out that the rejective attitudes of the mothers were associated with, and possibly caused by, an unhappy attitude toward marriage.

One of the most extensive studies of relations between the behavior of young children and characteristics of the home was published in 1936 by LaBerta Hattwick (8). Her data for the children's behavior comprised ratings based on several teachers' composite evaluations; ratings on home factors were based on the judgment of one teacher who referred to school conferences with parents and to a home visit. The analysis employed a large number of children drawn from 18 different nursery schools. Tetrachoric correlations were determined between a large number of different types of preschool behavior and 11 different ratings for the home. For the most part the ratings concerning the home were descriptive primarily of parental behavior which would be relevant for the child. The analyses show more promising correlations than could be explained plausibly on the basis of chance, and the behavior patterns implied by the major correlations are for the most part quite plausible. From numerous specific relationships she concludes that overattentive homes tend to produce children who display infantile and withdrawing reactions, and she shows that there is some tendency for neglectful homes to have children who are aggressive. In addition, there is indication that calm, happy homes produce children who are cooperative, and that impatient quarrelsome homes tend to be characterized by children who show uncooperative behavior.

Dorothy Baruch and J. Annie Wilcox describe an analysis of the relationships between ratings of the goodness of a child's adjustment (primarily interpersonal) and kinds of interparent tensions. This was in 1944 (3). The subjects comprised 76 preschool children and their parents. Intellectually the children and the parents were described as a superior group and participation in the study was dependent upon the parent's interest. The data were derived from observations of the child and interviews with the parents. One of the most important factors in the goodness of the children's adjustment was the absence of tension in the parents' sexual behavior; apparently, dissatisfaction concerning the sexual act indirectly contributed to an important degree to unsatisfactory adjustment in children. Lack of consideration of one spouse for another also contributed to unsatisfactory adjustment in the children, as did unsatisfactory expression of affection between spouses. Thus in general we see that inharmonious or incompatible parents appear to contribute to the adjustment problems of the children.

In 1945 Katherine Read published an account of an analysis which indicated that mothers who expressed approval of liberalism or freedom in the control of children had children whose behavior as measured by the Read-Conrad Inventory was relatively desirable (14). The mothers who approved of strict parental control had children whose behavior was indicated as unfavorable by the Read-Conrad Inventory. The mothers' attitudes were secured by means of a questionnaire whereas the children's behavior was rated by several nursery school teachers.

Although we shall generally avoid an employment of the case history material in our present discussion of the background for our problem, the implications of Watson's and of Read's study provide interesting corroboration in a clinical report of Edith Jackson published in 1947 (9). She describes three cases where strict discipline in early childhood was followed by the development of reactions that could be described as fearful, demanding, and aggressive.

Baldwin, Kalhorn, and Breese, in their 1945 monograph on the patterns of parent behavior (1), present interesting data which is consistent in its implications with trends reported by other investigators. They describe relationships between their parent behavior rating scales (used by social workers in visits to the home) and several characteristics of young school children. Among the numerous findings reported is an interesting relationship between parental rejection and nonconformist attitudes on the part of the children.

It is interesting to note that the tangible reports based upon systematic studies of practical behavior frequently draw the reader's attention to the importance of "democratic" home policies. "Reasonable" control policies appear to be important in assisting the child to become a well-socialized person who can live comfortably with others. Apparently arbitrariness on the part of the parents, or lack of sympathy and understanding, is conducive to the development of negative and destructive reactions on the part of the child.

In 1951 Barbara Bishop described a study based on 34 mothers and their preschool children (4). This report is based on systematic observations of mother-child interaction. The mothers were described with respect to such categories as lack of contact, directing, criticizing, and noncooperation. The child's behavior was categorized with respect to bid for attention, noncooperation, asking for information, and asking for help. Numerous correlations were reported which were not only statistically significant but which were often in the fifties and the sixties. Although these correlations can scarcely be taken as an evidence that long term patterns of child behavior are associated with important patterns of parental response, this report illustrates nicely that in many given situations an important correspondence exists between the behavior of mothers and the behavior of children.

An unpublished doctoral dissertation by Gerald Lesser describes a well-designed analysis based on (associates' judgment of) the behavior of 44 boys and the child-rearing practices of their mothers (as evaluated by an interview) (11). He reports that rejection by the mother is associated with overt aggression on the part of the child. He also finds that parents who advise aggressive activity have overtly aggressive children and that where aggression is strongly discouraged by the parents overt aggression on the part of the child is reduced.

Although not one of the foregoing references is likely to be regarded as compelling evidence for the importance of the environment in the development of the child's personality, each of them contributes to the plausibility of such a conclusion. For the most part, the samples were small and the measuring devices were not well standardized. In addition, there is little specific replication in content and methodology. Nevertheless, this stream of evidence continuing for the last 20 years has established an important pattern in our beliefs concerning the socialization of the individual. Some of the analyses, like those of Bishop (4), provide us with a kind of "laboratory exhibit" of the interactions which characterize the everyday existence of mother and child. This study obviously cannot be used as an indication of the long-term course of the child's development; it does illustrate however, that there is a systematic action and reaction, and they suggest descriptive components of the circumstances wherein interpersonal habits of the child may be formed in consequence of the pressures and requirements of the mother.

Despite their diversity, the studies of Watson (16), Newell (13), Hattwick (8), Baruch (3) Read (14), and Lesser (11) all suggest strongly that frustration of the child by neglect or harsh treatment evokes in him characteristic destructive responses and fearful attitudes.

Watson's study (16) indicates that strict parents may have more than their share of children whose conspicuous characteristics include aggression and anxiety. If relatively high scores on our Unsympathetic, Severe Toilet Training, and Punishment for Aggression cluster scores are indications of strictness on the part of the parents, it may be seen that the relations shown in Tables 1 and 2 are consistent with this part of Watson's study. Newell's study (13) seems to have similar implications and specifically indicates that hostile rejection on the part of the parents is associated with aggression on the part of the children. We find evidence of this in both our younger and older samples. Newell also points out that the rejective attitude of the mother may be caused by the unhappy attitude toward her marriage. We found (17) that rejective conditions in the home were inversely related with Harmonious Family Relationships and positively related with Parental Incompatability; this is true for both the younger and the older samples. All of this, incidentally, is in agreement with the implications of the study by Baruch and Wilcox (3). It is of interest to note that Baldwin, Kalhorn, and Breese (2) provide evidence of relationships between parental rejection and nonconformist attitudes on the part of the children. Lesser (11) also indicates that rejective mothers have aggressive children.

Consistencies between our report and other published reports could be shown in numerous detailed respects, but we are most interested in the broad implications of these investigations. In general, they strongly suggest that inharmonious, incompatible homes are likely to be rejective, and that such rejective homes are characterized by children who are fearful and aggressive. We believe that our study provides an important confirmation of the implications of these published studies. We have shown the pattern of relationships with respect to a wide constellation of variables, and the nature of our sample makes it seem particularly probable that such relationships (between aspects of the home and certain personal-social attributes of the children) as these are not a product of confounding with hereditary (i.e., we have no reason for believing in our sample that the particular relevant knowledge of the true parent and the particular relevant knowledge of the adoptive parent was consistently available to the agencies and systematically employed by them as a basis for selective placement of adoptive infants). It would seem, therefore, that if the agencies wish to reduce the possibility of creating adoptive homes where the child is fearful and aggressive, when they evaluate candidates they should be particularly interested in the harmony of the home, the compatibility of the spouses, and the improbability of the parents being rejective, even in times of disharmony or distress.

SUMMARY

We have explored relationships between various aspects of the children's development and various aspects of their environment for two samples. The younger sample comprised 114 children who were in their fifth or sixth year of life and not yet in the first grade. The older sample of 81 children was past the first grade and not yet pre-adolescent; 70 per cent of these children were in their eighth or ninth year of life. The measures of the children's development which we used included intelligence tests, scholastic achievement tests, measures of growth, and numerous facets of personal-social development. Among these facets of per-

sonal-social development are included the child's statement of his own behavior, the adoptive mother's description of the child's behavior, and examiners' ratings. Some of our measures of the child's development are tried and true; others are novel and untried.

Our measures of the environment were descriptive of the values and status of the adoptive home. Descriptions of the child-rearing practices of the adoptive parents were also available.

The results presented are in many respects similar for both the older and younger samples and lead us to believe that they are not the capricious result of sampling fluctuations.

In presenting the findings we have taken pains to draw the reader's attention to possible sources of confounding. Although the possibility that confounding could exist is no evidence that it did exist, we wish to remind the reader of the possibility, because our results have numerous practical implications and strongly suggest that differences in the environment may find expression in differences in the child's development, and this possibility seems particularly strong for the relationships where we can think of no plausible source of confounding which could detract from the environmental implications of our trends.

The relationships described by this study are not particularly strong and if, as we suspect, many are an expression of the formative role of the environmental differences, we would suggest that they should not be taken as indications of the possible maximal importance of these environmental differences. Because of obvious imperfections in our measures, we believe that the importance of the environment, predictive or causative, may be much greater than indicated by our correlations. For example, the correlation between phobic reactions in the child and unsympathetic child-rearing practices was .33 in our younger sample; neither of these variables is particularly reliable, however, and if such a relationship were examined by the use of reliable measure one might find a practically important relationship (when the well-known correction for attenuation is applied to this correlation of .33, it is raised to .66; when the correlation of .35 between unsympathetic child-rearing practices and anxious reactions in the older children is corrected for attenuation, an estimated correlation of .80 emerges).

Our study was designed as an exploration of possible relationships and was not committed to causal explanation of the relationships. Nevertheless, no plausible alternative explanations (e.g., confounding) are available for many of our relationships, and from these relationships certain important practical implications emerge. One implication is worthy of particular emphasis: inharmonious, incompatible, and rejective adoptive parents may tend to produce children who are aggressive and fearful.

GENERAL SUMMARY

The present report is the third in a series of three monographs: "Interviews as a source of scores for children and their homes" (17), "The predictive validity of the Yale Developmental Examination of Infant Behavior" (18), and "Relationships between some aspects of development and some aspects of environment for adoptive children." These monographs describe three successive phases of a program of research which was designed to provide answers for two questions:

1. How accurately and in what respects can the Yale Developmental Examination of infant behavior be employed to

predict future development for adoptive children?

2. What characteristics of adoptive homes (placements) may be shown to be correlated with characteristics of adoptive children?

Since the subjects for our investigation were adopted children and adoptive parents, certain limitations were imposed upon us. Adoptive families have a special status in our society and an investigation of the children and the child-rearing practices of such families involves special limitations and restrictions; it was necessary for us to plan an investigation which involved a minimum of threat and inconvenience for the cooperating adoptive families. In addition, it was necessary for us to devise procedures for evaluating adoptive children and their homes. However, the desirability of our providing special instruments for evaluation and measurement is not entirely due to the adoptive status of our subjects. There is a conspicuous dearth of practicable, time-tested procedures for evaluating the status of young children and for evaluating the values and the child-rearing practices of the home. Our efforts in this direction are described in the first section of this report.

Since our subjects were voluntary participants and since most of them were recruited with the help of the agencies who mediated the adoptive placements, there were some selective possibilities in our sample which had to be considered. The selective possibility that one might fear most is that the children included in our follow-up study were in some respects different as infants from those whom we were unable to include. Using the various parts of the infant examination as a basis for comparisons, we found no evidence for such a possibility. The indications available to us suggest that any differences between the families we saw and those we did not see were not due to the child per se but could represent the attitudes of the families toward the Yale Clinic of Child Development or toward the child-placement agencies.

In our exploration of the predictive validity of the infant examination (18), we used many different criteria of the child's development and all different aspects of the infant examination. Nevertheless, we failed to find evidence that the infant examination had a useful predictive validity. It should be emphasized that we examined the predictive validity of all the various facets of the infant examination and that we employed all feasible criteria of development for the adoptive children. Our criteria included the usual measures of mental and scholastic development, the various parts of a standard pediatric-neurological examination, and numerous scores and ratings descriptive of the child's personal and social development. These scores included the child's report on his own behavior, the adoptive mother's description of his behavior, and ratings made by the examiners.

Although we found no evidence for a useful degree of predictive validity and, therefore, cannot refute the hypothesis that the infant examination has no predictive validity, our results should not be taken as positive proof of this hypothesis. It is possible that, had we used some other criteria or had we used an older sample of children, some evidence of predictive validity could have emerged. Nevertheless, we did make a comprehensive effort to find evidence of predictive validity and we failed to find it. Other investigators using measures of mental ability alone have failed similarly to find evidence of predictive validity for infant examinations. It would seem, therefore, that those who strongly believe that the infant examination has predictive validity should now assume some responsibility for providing systematic data which indicate the nature, if not the degree, of such predictive value.

It should be remembered, however, that we have examined the predictive validity of an *infant* examination. All of the children who were included in our study had received an infant examination before they were 14 months of age. It should also be noted that the children who participated in our study were brought to the Yale Clinic of Child Development for evaluation in connection with their adoption or candidacy for adoption. Children who were grossly ill or obviously handicapped were seen at the Clinic in great numbers, but not as adoptive candidates. We have not sought to evaluate the use of the infant examination as a diagnostic aid for the evaluation of ill or handicapped children; our validative study is not concerned with the general value of the infant examination procedure as a device for evaluating the developmental status of an infant at any one time. We have been concerned with the predictive validity of the infant examination for infants who are possible candidates for adoption, and our results should not be generalized to other problems or situations, particularly of a medical nature, where the current evaluation or diagnosis of an infant is desired.

In this, our third monograph, are described some relationships between various aspects of the child's development and various aspects of his environment. There are descriptions of numerous interesting relationships, but our study, like others, provides no certain causal explanation for them. The analyses strongly suggest, however, that the child's environment may provide opportunities and limitations which result in differences in a child's development. This is not a novel indication, but our results are novel in that they provide evidence of such relationships in respects which have not heretofore been systematically described, and provide evidence for samples of children who are living in adoptive homes and are not biologically related to their parents. This suggests that some of the relationships between aspects of the child's development and aspects of his environment should be independent of an hereditary explanation, and therefore points definitely toward the importance of environmental differences as a factor in the development of personal-social differences in children.

Here we have considered various sources of confounding which might qualify the value of the environmental hypothesis in describing the relationships. It is acknowledged that some of the relationships could have arisen as a result of confounding with hereditary differences. Specifically, knowledge of some of the attainments of the true parents was available for many of the adoptive infants and we have evidence that this knowledge was used selectively in placing the infants in their adoptive homes. If such knowledge of the true parents as their educational-occupational status expresses hereditary differences, then it is possible that the correlated characteristics of the adoptive parent could show a relationship with aspects of the child's development, such as Binet IQ, as a result of confounding with heredity. We have no means of evaluating such an alternative hypothesis for some of the results. For many of our results, however, the relationship between characteristics of adoptive parents and characteristics of adoptive children could scarcely be described as a confounding with heredity.

In the first place, it is highly improbable that the necessary knowledge of the true parents' characteristics and of the environment to be provided by the adoptive applicants could have been known by the placement workers or could have had a role in the selective placement of the infants. In addition, it is improbable that the selective placement which would be required by our relationships could have occurred had such knowledge been available: the necessary beliefs concerning heredity are not sufficiently specific and we suspect that only *post hoc* hereditary hypotheses could be presented to describe many of our results.

In our study of the relationships between children's development and their environment, we have considered other sources of confounding, such as some hidden predictive value of the infant examination, and such as some systematic bias on the part of the mothers or on the part of our interviewers. For many of our relationships we can show that such sources of confounding are very implausible, indeed. It should be noted, however, that in those instances where the confounding hypothesis remains unchallenged, it cannot be claimed to be true, and that even for those possible instances one must consider the practical implications of the importance suggested by our data for the environmental differences.

It should be emphasized that our data provide numerous relationships for which no confounding hypothesis can be offered. For the most part, these relationships are suggestive of environmental sources or circumstances which influence the personal-social aspects of the child's development.

In general, it seems that our total study has two broadly practical implications. First, it offers no encouragement to those who would depend upon an *infant* examination as a basis for anticipating the subsequent development of adoptive children. Second, it strongly suggests that many aspects of the development of children, particularly those we might call personal or social, emerge at least in part as a result of differences in the child's environment, particularly those differences which have to do with the values of the home and its child-rearing practices. From these two general results, one might suspect (if one desires to maximize the possibility of certain patterns of development in adoptive children) that much of the effort commonly devoted to an exact evaluation of infants considered for adoption could more profitably be devoted to a study of the applicants who desire to be adoptive parents. With respect to the welfare of the adoptive children and adoptive families, our data suggest that additional research would provide the basis for some training of adoptive applicants and of adoptive parents in certain aspects of child rearing. This, however, is for the future. For the present it is hoped that our exploration will encourage others in their efforts to devise and validate sound instructional procedures for those who desire to create homes where children are friendly, confident, and have good ability.

REFERENCES

1. Baldwin, A. L., Kalhorn, Joan, & Breese, Fay H. Patterns of parent behavior. *Psychol. Monogr.*, 1945, **58**, No. 3 (Whole No. 268).
2. Baldwin, A. L., Kalhorn, Joan, & Breese, Fay H. The appraisal of parent behavior. *Psychol. Monogr.*, 1949, **63**, No. 4 (Whole No. 299).
3. Baruch, Dorothy W., & Wilcox, J. Annie. A study of sex differences in preschool children's adjustment consistent with interpa-

rental tensions. *J. genet. Psychol.*, 1944, **64**, 281-303.

4. BISHOP, BARBARA M. Mother-child interaction and the social behavior of children. *Psychol. Monogr.*, 1951, **65**, No. 11 (Whole No. 328).
5. BRENNER, RUTH F., *et al. A follow-up study of adoptive families.* New York: Child Adoption Research Committee, Inc., 1951.
6. BURKS, BARBARA S. The relative influence of nature and nurture upon mental development: a comparative study of foster parent-foster child resemblance and true parent-true child resemblance. *Yearb. Nat. Soc. Stud. Educ.*, Part I, 1928, **10**, 219-316.
7. FREEMAN, F. N., HOLZINGER, K. J., & MITCHELL, B. C. The influence of environment on the intelligence, school achievement, and conduct of foster children. *Yearb. Nat. Soc. Stud. Educ.*, Part I, 1928, **9**, 103-217.
8. HATTWICK, LABERTA. Interrelations between the preschool child's behavior and certain factors in the home. *Child Develpm.*, 1936, **7**, 200-226.
9. JACKSON, EDITH B. Clinical sidelights on learning and discipline. *Amer. J. Orthopsychiat.*, 1947, **17**, 580-589.
10. LEAHY, ALICE M. Nature-nurture and intelligence. *Genet. Psychol. Monogr.*, 1935, **17**, 236-308.
11. LESSER, G. Maternal attitudes and practices and the aggressive behavior of children. Unpublished doctors dissertation, Yale Univer., 1952.
12. LEVY, D. M. Relation of maternal overprotection to school grades and intelligence tests. *Amer J. Orthopsychiat.*, 1933, **3**, 26-34.
13. NEWELL, H. W. A further study of maternal rejection. *Amer J. Orthopsychiat.*, 1936, **6**, 576-589.
14. READ, KATHERINE H. Parents' expressed attitudes and children's behavior. *J. consult. Psychol.*, 1945, **9**, 95-100.
15. SIMS, V. M. The influence of blood relationship and common environment on measured intelligence. *J. educ. Psychol.*, 1931, **22**, 56-65.
16. WATSON, G. A comparison of the effects of lax versus strict home training. *J. soc. Psychol.*, 1934, **5**, 102-105.
17. WITTENBORN, J. R. Interviews as a source of scores for children and their homes. *Psychol. Monogr.*, 1956, **70**, No. 1 (Whole No. 408).
18. WITTENBORN, J. R. The predictive validity of the Yale developmental examination of infant behavior. *Psychol. Monogr.*, 1956, **70**, No. 2 (Whole No. 409).

(Accepted for early publication October 18, 1955)

Vol. 70, No. 4 Whole No. 411, 1956

Psychological Monographs: General and Applied

The Tendency to See Objects As Equidistant and Its Inverse Relation to Lateral Separation[1]

Walter C. Gogel
Army Medical Research Laboratory[2]

INTRODUCTION

WHEN distant objects are viewed without optical devices, or when objects are viewed through a device which uses only one eye, binocular disparity (stereopsis) becomes weak or inoperative in determining distance perceptions. If, in addition, the objects appear in a relatively homogeneous field of view, such as an expanse of sky or snow, many of the perspective and overlay cues of distance perception are eliminated. It is of some importance to determine the errors which would occur in the perception of the relative distance of objects under such restricted visual conditions. Even in a situation in which the distance aspects are generally well structured, there may be some objects which have few indicators to relate their apparent distance to that of other objects.

Consider the case in which there are *no* cues present to determine the apparent relative distances of objects. Suppose that under these conditions objects are physically located at different distances from an observer. Where will the objects be seen in depth relative to each other? One possible answer would be that the objects would appear indefinitely located in relative distance. This answer assumes that there is no tendency to see the objects in any particular distance relation when the distance cues are removed. If this answer is correct, it might be expected that, in a situation in which the cues to the relative distance of objects were *severely limited,* the perceived relative distance of the objects would be in agreement with whatever distance cues remained; i.e., the observer would correctly use whatever cues were available in structuring the apparent depth aspects of the visual field. Another answer would be that, in the complete absence of distance cues between objects, all objects would appear to be at the same distance from the observer. If this second answer is correct, a different result would be expected in the situation in which the distance cues between objects were very limited. The tendency to see the objects as equidistant would be opposed to the weak distance cues, with the result that the physically nonequidistant objects would appear to have less depth difference than is actually the case. In the continuum from a perceptually well-defined depth situation to one of very limited cues, it would be expected, if the second answer is correct,

---

[1] The author wishes to thank John P. Tammaro and Kay Inaba for their help in collecting and analyzing the data.

[2] The opinions or conclusions contained in the present report are those of the author. They are not to be construed as reflecting the views or endorsement of the Department of the Army.

Bureau Ednl. Psy. Research
DAVID HARE TRAINING COLLEGE
Dated
Acc. No

that the tendency to see objects as equidistant would become increasingly important.

There is some evidence that the second is the more accurate answer. Judd (4) has reported that a series of white threads presented against a black background and located at different distances from an observer seemed to be at the same distances as each other when viewed monocularly; whereas, when viewed binocularly, their differences in distance were clearly perceived. In discussing the apparent depth between double images and a binocularly fixated object, Judd states, "And we may assert it as a general principle that in monocular vision objects tend toward a single plane, the plane determined by binocular factors." (4, p. 293.)

In the present study monocular and binocular objects were viewed simultaneously. It is assumed that, under the conditions of this study, stereopsis was effectively absent between a monocular and any of the binocular objects.[3] This situation, in the absence of other depth cues, was used to investigate the question of where two physically nonequidistant objects would appear in distance when cues to their relative distance are weak or absent. From the observations of Judd, it might be expected that a monocular object would tend to appear at the same distance from an observer as a binocular object when there are no other binocular objects in the field of view and when monocular indicators of the relative depth between the two objects are weak or absent. This was studied quantitatively in the first experiment of the present study.

What would the perception be, however, if more than one binocular object were present, with the several binocular objects located at different distances from the observer? Where, in depth, would the monocular object appear under these conditions? The depth between the binocularly viewed objects would be fairly accurately perceived because of the binocular disparity cues. But the monocular object can be arranged to present few or no indications of its relative depth position. This is the type of situation in which the apparent relative distances of some, but not all, of the objects are well defined. This was investigated in the second experiment of this study.

In the third and fourth experiments of this study, size cues were introduced between the monocular object and one or all of the binocular objects. The fourth experiment also was designed to "explain" the results from an experiment discussed in a previous study (3), in which it was found that the effectiveness of the size cue in determining the apparent depth between physically equidistant playing cards, one monocular and one binocular, increased as the lateral separation of the playing cards was increased.

GENERAL METHOD AND APPARATUS

A method of determining the apparent depth between the monocular and binocular object or objects was required. The method used in three out of four of the present experiments was that described previously (3) in which a small binocular disc was adjusted to apparent distance equality with each of several objects in the field of view. The subject adjusted this disc in its distance from himself but he did not adjust the direction (line of sight) of the disc from himself. This direction was set by the experimenter so that the line of sight to the disc was always close to the direction (line of sight) of one of the binocular objects. Consistent with this method, this line-of-sight position of the disc was always closest to that of the same binocular object, regardless of the object to which the disc was to be adjusted in distance. Under these conditions, the difference between the distance adjustment of the disc to apparent equidistance with each of the several objects was taken as a measure of the apparent depth between these objects. It is as though the stereopsis between the binocular disc and its closest line-of-sight binocular object was used in adjusting the disc to apparent equidistance with any other object in the field of view, regardless of whether these other objects were monocularly or binocularly viewed. The experimental justification for this method of measuring apparent distance is contained in a previous report (2).

---

[3] That this is not true under all conditions is illustrated by the Wheatstone-Panum limiting case (5, p. 181-182).

The disc used in this study was stereoscopically generated rather than being an actual object in the field of view. The particular stereoscopic instrument used to generate the disc has been described by Alluisi and Harker (1). It is the same instrument which was employed in the previous study (3). This instrument consisted of two optical systems, one before each eye of the subject. The right and left optical system independently generated a source of parallel light which was projected as a disc image on the right and left retina, respectively. The right optical system could be rotated by turning an adjustment knob. This changed in continuous amounts the direction at which the parallel light entered the right eye of the subject. The left optical system could not be rotated. Each optical system contained a reflecting-transmitting surface which was positioned so that both the disc image and actual objects in the field of view could be simultaneously visible. The subject looked through a pair of restrictive apertures and binocularly fused the retinal disc images. This resulted in the disc appearing to be located in distance with respect to an actual object. By turning the adjustment knob, the subject could make the disc appear to move in distance with respect to the actual object, and the disc could be adjusted to apparent equidistance with this object. From the interpupillary distance of the subject and the calibration constant of the instrument, the distance position of the disc in centimeters could be calculated for any position at which the subject adjusted the disc. For the purposes of this study, the disc can be regarded as essentially a real object which actually moved toward or away from the subject, depending upon the direction in which the subject turned the adjustment knob. The disc was always viewed binocularly. It was orange in color and subtended 15′ of visual angle.

It usually occurred that one of the objects which was physically located in front of the subject was to be seen monocularly. When this was required, a small, black, opaque screen was introduced between the object and one eye of the subject. The object was visible, however, to the other eye of the subject. The room was dark except for the objects used in the experiment, and the screen was not visible. This screen was carefully placed so that, when it was positioned to make one of the objects monocular, it did not interfere with the binocular view of the other object or objects.

The subjects for these experiments were drawn from a group of approximately 24 men, who were experienced in using the stereoscopic instrument.

EXPERIMENT I

Display

Schematic top-view drawings of the situations presented to the subjects are shown in Fig. 1. Object S was a 4.5-cm. white square. Object R was a white rectangle, 2.5 cm. wide and 8.0 cm. high. The centers of the square and rectangle were the same height from the floor. The square was always placed so that its right edge was 1° 53′ of visual angle to the left of the left edge of the rectangle. The square was always viewed binocularly. The rectangle was seen either binocularly or with only the left eye of the subject. The rectangle was always located 303 cm. from the subject. The square was either 250 cm. from the subject (Part A of Fig. 1), or 340 cm. from the subject (Part B of Fig. 1). This resulted in a total of four situations. They were as follows: With the square observed binocularly and the rectangle observed monocularly, the square was (*a*) physically closer to the subject than the rectangle, or (*b*) physically farther from the subject than the rectangle. With both the square and rectangle observed binocularly, the square was (*c*) physically closer to the subject than the rectangle, or (*d*) physically farther from the subject than the rectangle.

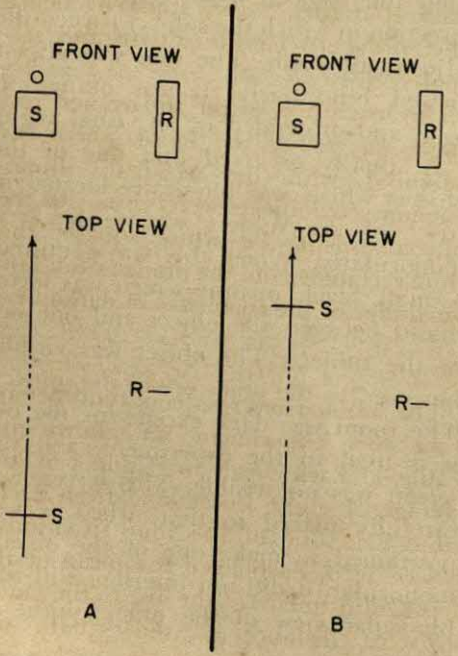


FIG. 1. Schematic front and top-view drawings of the situations used to investigate the apparent depth position of a monocular or binocular rectangle (R) as a function of the depth position of a binocular square (S).

Part A of Fig. 1 illustrates Situations *a* and *c*, and Part B of Fig. 1 illustrates Situations *b* and *d*. It will be observed that the only difference between Parts A and B of Fig. 1 was that the square was moved from a position in front of the rectangle to a position behind the rectangle. With Situation *a* and *b*, stereopsis was effectively absent between the square and rectangle, while with Situations *c* and *d*, the stereopsis cue was present between these two objects. The main interest of the experiment is with the situations in which the rectangle was viewed monocularly (Situations *a* and *b*). The purpose of Situations *c* and *d* is merely to demonstrate that the depth difference between the square and rectangle was readily perceived when the stereopsis cue occurred between the square and rectangle.

The long, arrowed lines in the top views of Fig. 1 approximate the apparent path of depth movement of the binocularly viewed disc in the depth vicinity of the square and rectangle when the subject turned the stereoscopic adjustment knob. If the position of the subject in the top-view drawings were shown in Fig. 1, it would be indicated below the top-view drawings. For all points on the long, arrowed lines, the line-of-sight position of the disc was always closer to the line-of-sight position of the square than to the line-of-sight position of the rectangle. As indicated by the circle (representing the disc) in the front views of Fig. 1, the line-of-sight position of the disc was always above the square (by 10′ of visual angle). The same line-of-sight position of the disc was maintained regardless of whether the disc was adjusted to apparent depth equality with the square or with the rectangle.

Procedure

The experimenter adjusted the line-of-sight position of the disc so that this position was 10′ of visual angle above the square (see Fig. 1). This line-of-sight position was unchanged throughout the experiment. It was not modified by the subject, whose task was to adjust the stereoscopic instrument knob until the binocular disc appeared to be at the same distance from himself as either the square or rectangle. The difference between the average adjustment of the disc to apparent distance equality with the square and with the rectangle is taken as a measure of the apparent depth between the square and rectangle (3). For any presentation of the square and rectangle, 16 adjustments of the disc to the same apparent distance as one of these objects was followed by 16 adjustments to apparent equidistance with the other object. A bracketing technique was used. In each trial the subject adjusted the disc to appear alternately farther and nearer than the square (or rectangle), with the final approach to the same apparent depth position as the square (or rectangle) occurring with the disc appearing to move toward or away from the subject, according to the instructions. With half the subjects, the final approach to apparent depth equality in the first eight of these depth adjustments was made with the disc moving toward the subject, and in the next eight adjustments with the disc moving away from him. With the remaining subjects, the order was reversed. Eight subjects were used. Each of these adjusted the disc to apparent distance equality with each of the two objects (the square and the rectangle) in each of the four situations, *a, b, c,* and *d.* The procedure was systematically varied with respect to (*a*) whether the square or rectangle was designated first as the object to which the disc was adjusted to apparent equidistance, (*b*) whether the square was presented first behind or first in front of the rectangle, and (*c*) whether the rectangle was presented first monocularly or first binocularly.

After the 16 adjustments of the disc to the same apparent distance as the rectangle and the square in a particular situation were completed, the subject was asked to estimate the apparent depth difference between the square and rectangle.

The brightness of the square and rectangle was always 2.4 foot-lamberts, as measured with the Macbeth illuminometer. No objects were visible to the subject except the white square, the white rectangle, and the orange disc. The remainder of the room was in darkness.

Results

The summarized results in centimeters from this experiment are shown in Table 1. Each mean in Table 1 is an average of eight scores, one from each subject, in one of the four situations, *a, b, c,* or *d.* Each score is a mean of 16 depth adjustments of the disc to the same apparent distance as the square or rectangle. The standard deviations of Table 1 were computed from the distributions of eight scores. The statistical significance of the differences between the various means was determined by the

TABLE 1

Means and Standard Deviations in Centimeters of the Adjustment of a Binocular Disc to Apparent Equidistance with a Monocular or Binocular Rectangle (R) and a Binocular Square (S), as a Function of the Depth Position of the Square

Distance adjustment of binocular disc with respect to:	R Monocular				R Binocular			
	S in Front of R (Situation *a*)		S Behind R (Situation *b*)		S in Front of R (Situation *c*)		S Behind R (Situation *d*)	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
S	254	4.7	342	6.1	254	4.1	343	6.1
R	257	11.7	337	11.4	302	8.8	310	8.6

t test. In calculating the *t* values (in this and the following experiments), a distribution of difference scores was formed from the distributions whose mean difference was to be tested for significance.

Consider the results in the left half of Table 1, obtained from the situations in which the rectangle (R) was viewed monocularly. When the binocular square (S) was located at a distance of 250 cm. from the subject and the rectangle at a distance of 303 cm. (Situation *a*), the average distance adjustment of the binocular disc with respect to the square was 254 cm., and with respect to the rectangle 257 cm. This difference of 3 cm. was not significant at the .10 level of confidence ($t = .7$). This suggests that there was little or no difference in apparent depth between the square and rectangle. Also, when the binocular square was located at 340 cm., with the physical position of the monocular rectangle still at 303 cm. (Situation *b*), the average apparent depth between the rectangle and square was 5 cm. (342 cm. minus 337 cm.). This difference also was not significant at the .10 level of confidence ($t = 1.4$). From these results it seems that the monocular rectangle appeared to be located near or at the distance position of the binocular square for each of the two distance positions of the square. If there was a tendency for the monocular rectangle to appear at the distance of the binocular square, the mean adjustment of the disc to apparent equidistance with the rectangle should differ in Situations *a* and *b*, with Situation *b* resulting in the larger centimeter value. The mean adjustment results on the rectangle were 337 cm. and 257 cm. from Situations *b* and *a*, respectively. This difference of 80 cm. is significant beyond the .001 level of confidence ($t = 11.6$). These results support the conclusion that there is a tendency to see a monocular object at the same apparent distance as a binocular object.

The right half of Table 1 refers to Situations *c* and *d*, in which both the square and rectangle were observed binocularly. It will be seen from these results that the rectangle (R) was clearly (and correctly) perceived as being behind the square (S) in Situation *c*, and in front of the square in Situation *d*. The perceived differences in distance (302 cm. minus 254 cm., and 343 cm. minus 310 cm.) were each significant beyond the .001 level of confidence (*t* being 17.0 and 17.1, respectively). The difference between the mean adjustment results on the rectangle from Situations *d* and *c* was 8 cm. (310 cm. minus 302 cm.). This difference was between the .10 and .05 levels of confidence ($t = 2.0$). One possible interpretation of this 8-cm. difference is that there may have been some tendency to see the rectangle and square as less different in depth than they actually were. But this tendency was much less effective in the situations in which the rectangle was viewed binocularly than in the situations in which the rectangle was viewed monocularly.

The average verbal report when the square was physically less distant than the rectangle was that the square was 3 in. in front of the rectangle when the rectangle was observed monocularly, and 33 in. in front of the rectangle when the rectangle was observed binocularly. The average verbal report when the square was physically more distant than the rectangle was that the square was 11 in. behind the rectangle when the rectangle was observed monocularly, and 26 in. behind the rectangle when the rectangle was observed binocularly. The direction of these changes in the average verbal report with the change in the method of observing the rectangle (monocular or binocular observation) is in agreement with the direction of the corresponding changes in the mean differences of Table 1.

Discussion

This experiment demonstrates quantitatively that there is a tendency to see a monocular object at the same apparent distance as a binocular object. Actually, there were three, not two, objects in view: the square, the rectangle, and the disc. The only reason for having the disc present was to determine the apparent depth difference between the square and rectangle. Consider the two Situations (*a* and *b*) in which the rectangle was observed monocularly. If the monocular rectangle tended to be seen at the same apparent depth as a binocular object, the presence of the binocular disc might be expected to affect the apparent depth position of the monocular rectangle, whenever the disc was at an apparent depth position which was different from that of the binocular square. But the disc was present and had the same line-of-sight position, both in the situation in which the binocular square was behind the monocular rectangle and in the situation in which the binocular square was in front of the monocular rectangle. Therefore, the difference between the 337- and 257-cm. average depth adjustment of the disc to apparent equidistance with the monocular rectangle can only be attributed to the factor which changed between the two situations, i.e., to the change in the distance position of the binocular square.

EXPERIMENT II

Display

In Experiment I it was quantitatively demonstrated that there is a tendency to see a monocular and binocular object at the same apparent distance. The purpose of Experiment II was to determine whether the strength of this tendency is a function of the lateral line-of-sight separation of the objects. The words *lateral separation,* as they are used here and in the following experiments, refer to relative right or left positions, i.e., to the difference in visual right or left direction between the several objects. Two or more objects can be made to assume any lateral separation with respect to each other, while remaining different in their distances from the subject. In the present experiment, two binocularly viewed white squares (4.5 cm. on a side) and a monocularly viewed white rectangle (2.5 cm. wide and 8.0 cm. high) were used. The monocular object here, as in Experiment I, was of a different shape than the binocular object or objects in order to minimize the use of any size cue between the monocular object and

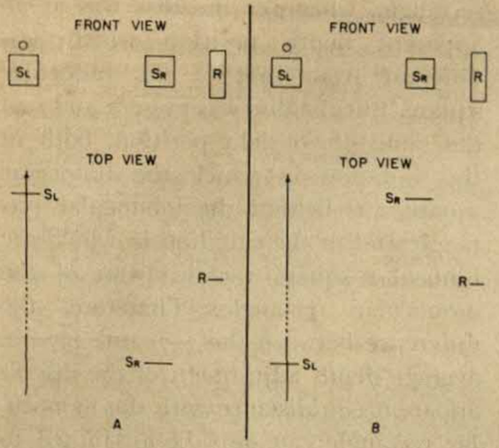


FIG. 2. Schematic front and top-view drawings of the situations used to investigate the apparent depth position of a monocular rectangle (R) as a function of the depth and lateral positions of two binocular squares (S_R and S_L).

a binocular object. Between the binocular squares, however, it was appropriate to have the usual size cue resulting from using similar objects, since it was necessary that the relative distance positions of the squares with respect to each other be clearly perceived.

The two binocular squares and the monocular rectangle were arranged as illustrated by Fig. 2. In Part A of Fig. 2, the square on the left (S_L) was 416 cm. from the subject, and the square on the right (S_R) was 346 cm. from the subject. In Part B of Fig. 2, the distance positions of the two squares were reversed. In this case, the square on the left (S_L) was 346 cm. from the subject and the square on the right (S_R) was 416 cm. from the subject. The monocular rectangle (R) was 381 cm. from the subject in both Part A and Part B of Fig. 2, i.e., physically, the monocular rectangle was always midway in depth between the two binocular squares. While the depth positions of the two squares were reversed in the situations represented by the two parts of Fig. 2, the lateral line-of-sight positions of the squares remained unchanged. In both Part A and Part B of Fig. 2, S_L was 3° 49′ of visual angle to the left of the monocular rectangle (R), while S_R was only 46′ to the left of R. The rectangle was seen always with only the left eye, while the two squares and the disc were always seen with both eyes, i.e., the stereopsis cue occurred between the two squares and between the disc and each square, but was effectively absent between the rectangle and either of the squares or the disc.

The angle formed between the lines-of-sight to R and S_R was considerably less than that between R and S_L. The rectangle (R) was laterally closer to S_R than to S_L. If R, which was physically midway in depth between S_R and S_L, appeared to be located in depth closer to S_R than to S_L, this would be evidence that the strength of the tendency to see a monocular object as equidistant with a binocular object is inversely related to lateral directional separation.

The long, arrowed lines in the top views of Fig. 2 approximate the apparent path of depth movement of the disc in the depth vicinity of the squares and rectangle when the subject turned the adjustment knob. The disc was always seen with both eyes. The line-of-sight position of the disc was always 10′ of visual angle directly above the left square, regardless of where the disc appeared in depth. This position of the disc is indicated by the circle in the front views of Fig. 2. The disc retained this same line-of-sight orientation, regardless of the object to which it was adjusted in apparent depth.

The brightness of the two white squares and the white rectangle was always 2.4 foot-lamberts, as measured with a Macbeth Illuminometer. No objects were visible in the field of view except the two squares, the rectangle, and the disc.

Procedure

Twelve subjects were used. The task of the subject was to adjust the disc in depth until it appeared to be at the same distance as one of the objects in the field of view. Each of the two squares and the rectangle, in turn, was specified as the object to which the disc was to be adjusted in depth. Each subject was presented with the situation represented by both A and B of Fig. 2. Sixteen successive adjustments of the disc to each object were made by each subject. As in Experiment I, the bracketing technique was used, with the final distance adjustment of the disc systematically varied with respect to whether the disc moved toward or away from the subject. The procedure was systematically varied between subjects with respect to the order of designating each square and the rectangle as the object to which the disc was adjusted in depth, and with respect to whether the situation first presented to the subject was that represented by Part A or Part B of Fig. 2.

After completing the distance adjustments of the disc in either the situation represented by Part A or Part B of Fig. 2, the subject was asked to estimate the depth position of the rectangle with respect to each of the squares.

Results

The summarized results from Experiment II, in centimeters, are given in Table 2. The columns labeled "Mean," in Table 2, contain averages of 12 scores, one score from each subject, where each score is a mean of 16 adjustments of the disc to the same apparent depth as a square or the rectangle, as indicated by the left column. The standard deviations of Table 2 were computed from the distributions of twelve scores.

TABLE 2

MEANS AND STANDARD DEVIATIONS IN CENTIMETERS OF THE ADJUSTMENT OF A BINOCULAR DISC TO APPARENT EQUIDISTANCE WITH A MONOCULAR RECTANGLE (R) AND TWO BINOCULAR SQUARES (S_L AND S_R), AS A FUNCTION OF THE DEPTH AND LATERAL POSITIONS OF THE SQUARES

Distance adjustment of binocular disc with respect to:	S_L Far (A of Fig. 2)		S_L Near (B of Fig. 2)	
	Mean	*SD*	Mean	*SD*
S_L	414	7.4	344	6.4
R	368	23.9	398	36.8
S_R	350	10.7	415	10.7

The column labeled "S_L Far," in Table 2, refers to the situation represented by Part A of Fig. 2. The average apparent distance difference between the left square (S_L) and the rectangle R, as measured by using the disc, is 46 cm. 414 cm. minus 368 cm.). The average apparent distance difference between the right square (S_R) and the rectangle (R) is 18 cm. (368 cm. minus 350 cm.). This suggests that the monocular rectangle usually appeared to be closer in depth to the square to which it was laterally closer (S_R) than to the square from which it was laterally more displaced (S_L). Subtracting the 18-cm. difference from the 46-cm. difference gives 28 cm., which is significantly different from zero at the .05 level of confidence ($t = 2.2$). The column labeled "S_L Near," in Table 2, refers to the situation represented by Part B of Fig. 2. Here, the average apparent relative depth position of the rectangle, as measured by using the disc, is 54 cm. (398 cm. minus 344 cm.) in back of S_L, and 17 cm. (415 cm. minus 398 cm.) in front of S_R. The 54-cm difference is greater than the 17-cm. difference at the .10 level of confidence ($t = 1.8$).

Therefore, from the results of the situations represented by both Part A and Part B of Fig. 2, there is some evidence that the monocular rectangle appeared to be closer in depth to the square to which it was laterally closer (S_R) than to the square from which it was laterally more displaced (S_L). In one situation the rectangle appeared to be closer to the subject than its physical relative position, and in the other situation farther from the subject than its physical relative position. The physical position of the rectangle was the same in both situations. An over-all test of the change in the average apparent position of the rectangle between the two situations can be made. This difference of 30 cm. (398 cm. minus 368 cm.) is significantly different from zero beyond the .01 level of confidence ($t = 3.7$).

The disc adjustment difference of 18 cm. between R and S_R in the situation represented by Part A of Fig. 2 is significant beyond the .02 level of confidence ($t = 2.8$), but the adjustment difference of 17 cm. between R and S_R in the situation represented by Part B of Fig. 2 is not significant at the .10 level of confidence ($t = 1.7$). The 18-cm. adjustment difference suggests that, although the monocular rectangle tended to appear closer in depth to S_R than to S_L, it did not, on the average, appear at the distance of S_R. While the tendency to see R at the depth of S_L was less strong than the tendency to see R at the depth of S_R, the presence of S_L was sufficient to keep the rectangle from appearing at the distance of S_R.

The average verbal report for the situation represented by Part A of Fig. 2 was that the rectangle was 4 in. behind the right square and 25 inches in front of the left square. The average verbal report for the situation represented by Part B of Fig. 2 was that the rectangle was 2 in. in front of the right square and 34 in. behind the left square. These results are in general agreement with the results obtained from using the disc.

The evidence from this experiment is that, in a situation in which two binocular objects are present, a monocular object will tend to appear closest in depth to that binocular object which has the least lateral visual (line-of-sight) separation from the monocular object. It may be concluded that the strength of the tendency to see a monocular object at the same apparent depth as a binocular object decreases as the lateral visual separation of the two objects is increased. The results of this experiment also suggest that binocular objects, other than the laterally most adjacent object, had some effect upon the perceived distance position of the rectangle.

Discussion

Here, as in Experiment I, the disc constituted an additional binocular object. However, the presence of the disc cannot be used to explain the difference which occurred in the average adjustment of the disc to apparent equidistance with the rectangle between the situations represented by Part A and by Part B of Fig. 2. In both situations, the disc was not only present but had the same line-of-sight orientation with respect to the other objects. Therefore, the difference between the 398-cm. and 368-cm. mean adjustment can be attributed only to the change in the distance positions of the binocular squares. In fact, since the bracketing technique was used and the final approach of the disc to apparent distance equality with a particular object was varied with respect to whether the disc moved toward or away from the subject, it is rather difficult to see how

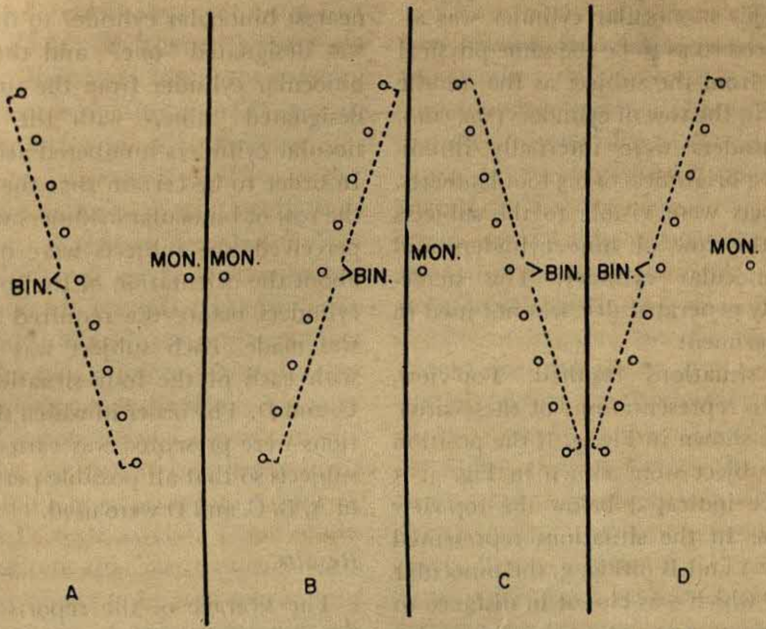


Fig. 3. Schematic top-view drawings of the situations used to investigate the apparent depth position of a monocular cylinder as a function of its lateral position with respect to nine binocular cylinders.

the presence of the disc could have systematically determined the apparent depth between the various objects, either in this experiment or in Experiment I.

The standard deviations of Table 2 with respect to the rectangle are large. It seems that there were some large, individual differences in the perception of the relative depth positions of the monocular rectangle.

Experiment III

Display

In this experiment, nine binocular objects were used. These consisted of a group of upright, internally illuminated white paper cylinders which were placed in a row, with the row located so that it formed a 17.5° angle with an imaginary line drawn from the subject through the middle cylinder. All the cylinders were 4.6 cm. in height and 1.7 cm. in diameter. The separation between adjacent cylinders in the row of cylinders was 15.2 cm.

The row of cylinders receded in distance either from near-right to far-left or from near-left to far-right. The middle cylinder was always 305 cm. from the subject. All the cylinders in the row of nine cylinders were always seen binocularly.

An additional cylinder of the same size was placed to the right or to the left of the row of nine cylinders. This additional cylinder was always seen monocularly. When the monocular cylinder was on the right it was seen with only the left eye, and when the monocular cylinder was on the left it was seen with only the right eye. When this monocular cylinder was on the right or the left, it was laterally separated (in terms of its line of sight) from the most right or most left cylinder in the row of binocular cylinders by 50′ of visual angle. This monocular cylinder was always placed so as to be the same physical distance from the subject as the middle cylinder in the row of cylinders (305 cm.). The cylinders were internally illuminated to a brightness of 0.3 foot-lamberts. No objects were visible to the subjects except the row of nine cylinders and the monocular cylinder. The stereoscopically generated disc was not used in this experiment.

Four situations resulted. Top-view, schematic representations of these situations are shown in Fig. 3. If the position of the subject were shown in Fig. 3, it would be indicated below the top-view drawings. In the situations represented by Parts A and B of Fig. 3, the binocular cylinder which was closest in distance to the subject was also the binocular cylinder which was laterally closest to the line of sight from the subject to the monocular cylinder. In the situations represented by C and D of Fig. 3, the binocular cylinder which was farthest in distance from the subject was also the binocular cylinder which was laterally closest to the line of sight from the subject to the monocular cylinder. If the inverse relation between lateral separation and the strength of the tendency to see a monocular object at the same distance as a binocular object was effective in these situations, it would be expected that the binocular cylinder which appeared to be at the same depth as the monocular cylinder would be a cylinder which was closer to the subject in Situations A and B than in Situations C and D of Fig. 3.

Procedure

Twenty-four subjects were used. Their task was to judge which of the cylinders in the row of (binocular) cylinders was most nearly at the same depth as the additional (monocular) cylinder. The nearest binocular cylinder to the subject was designated "one," and the farthest binocular cylinder from the subject was designated "nine," with the other binocular cylinders numbered accordingly. In order to be certain that the depth in the row of binocular cylinders was clearly perceived, the subjects were questioned about the orientation of the row of nine cylinders before the required judgment was made. Each subject was presented with each of the four situations, A, B, C, and D. The order in which these situations were presented was varied between subjects so that all possible permutations of A, B, C, and D were used.

Results

The average of the reports from the 24 subjects and the standard deviations of these reports are given in Table 3. The situations, A, B, C, and D, which are shown in the column headings of Table 3, are those represented in Fig. 3. For example, with Situation A of Fig. 3, the average of the reports was that the additional cylinder (the monocular cylinder) was at a depth position midway between the third and fourth cylinder (3.5), counting from the near end of the row of cylinders. In both Situations A and B, the monocular cylinder was, on the average, seen as closer in depth to the front cylinders in the row of cylinders than to the back cylinders, while in

TABLE 3

MEANS AND STANDARD DEVIATIONS OF THE APPARENT DEPTH POSITION OF A MONOCULAR CYLINDER, AS A FUNCTION OF ITS LATERAL POSITION WITH REGARD TO NINE BINOCULAR CYLINDERS*

Measure	Situation A	Situation B	Situation C	Situation D
Mean	3.5	3.4	7.3	7.4
SD	2.3	1.5	1.8	1.6

* For explanation of table, see text.

Situations C and D the reverse was true. The average report from A (3.5) was significantly different from the average report from either C (7.3) or D (7.4) beyond the .001 level of confidence ($t =$ 6.0 and 7.9, respectively). The average report from B (3.4) was significantly different from that from either C or D beyond the .001 level of confidence ($t = 7.5$ and 8.6, respectively). Also, each mean report differs from 5 (the number of the binocular cylinder whose distance from the subject was physically the same as that of the monocular object) beyond at least the .01 level of confidence (the t values being 3.0, 5.2, 6.2, and 7.0 for A, B, C, and D, respectively). These results clearly indicate that the tendency for a monocular object to be seen more in the depth direction of those binocular objects which were less laterally displaced from it was effective in these situations.

Discussion

Since all the cylinders were physically identical, size cues were available in the situations used in this experiment. It will be noted that the retinal size of the monocular cylinder was the same as that of the fifth cylinder. The results that were obtained occurred in spite of this size cue. It seems that the tendency for a monocular object to be seen at the depth of a binocular object should not be regarded as a factor which is important only when no other factors are available to determine the perception. It may, therefore, be of consequence in a variety of situations. It is suggested that this tendency is present to some extent in all situations in which more than one object is visible. But if other factors that localize a particular object in depth are strong, the effectiveness of this tendency on that particular object will be restricted.

This tendency and its relation to lateral separation should not be considered as operating only between a monocular and binocular object. There appears to be little reason why it should not also apply to the apparent depth between several monocular objects, or even between several binocular objects if, in this latter case, the binocular cues are sufficiently weak.[4]

The tendency to see objects as equidistant does not always act to eliminate the perception of depth differences. If depth occurs in the visual field through the action of other factors, the tendency to see objects as equidistant with its inverse relation to lateral separation may result in an increase in the apparent depth differences between some of the objects. For example, this factor produced an apparent depth difference in the present experiment between the monocular cylinder and the fifth cylinder in the row of binocular cylinders.

EXPERIMENT IV

Display

It was found in a previous study (3) that the apparent depth between two differently sized playing cards which were physically at the same distance from the subjects increased as the lateral separation of the cards was increased. The situation which is of interest here is the case in which one of the two cards was viewed binocularly and the other monocularly. Since stereopsis was effectively absent between the monocular card and the binocular card, the factors to be considered are the size cue and the tendency to see a monocular object at the same distance as a binocular object. In this particular case, these two factors were opposed in their effects, i.e., the size cue should have operated to make the larger card appear closer than the smaller card, while the above tendency should have operated to make the cards appear equidistant. It is possible, therefore, that the increase in the apparent depth between the cards when the lateral separation of the cards increased was due to the inverse relation between lateral separation and the tendency to see a monocular and binocular object as equidistant, rather than to any change in the strength of the size cue to relative depth.

---

[4] Applications of this tendency to situations of these types are found in Reports No. 148 and 157, Army Medical Research Laboratory, Fort Knox, Kentucky.

To investigate this, a binocular normal-sized playing card and a monocular double-sized playing card were used with a 22.9-cm. lateral separation between their inner edges. The double-sized card was on the right and the normal-sized card was on the left. Each of these two cards was located at a distance of 303 cm. from the subjects. This was identical with one of the situations used in the previous study (3), except that a diamond-shaped figure (11.9 cm. by 4.8 cm.) was sometimes present. When the diamond-shaped figure was used, it was placed 55 cm. behind the two physically equidistant playing cards, with the right corner of the diamond laterally displaced to the left of the double-sized playing card by 51 minutes of visual angle, and its left corner laterally displayed to the right of the normal-sized playing card by 2° 23′ of visual angle. Schematic representations of the situations used are shown in Fig. 4. Part *a* of Fig. 4 represents the situation in which the diamond was absent, and Part *b* of Fig. 4 the situation in which the diamond was present. The two situations are identical in all other respects. In both situations the size difference between the two cards would tend to make the double-sized card (Card B of Fig. 4) appear closer to the subject than the normal-sized card (Card A of Fig. 4). With the diamond present, the binocular disparity cue between the normal-sized card and the diamond should make the diamond appear to be more distant than this card. But there was effectively no stereopsis between the binocular diamond and the monocular double-sized card. Therefore, when the diamond was present, the tendency for the monocular object (the double-sized card) to be seen nearest in distance to the laterally nearer object (the diamond) would tend to make the double-sized card appear behind the normal-sized card. It would be expected, therefore, that when the diamond was present, the double-sized card should appear less far in front of the normal-sized card than in the situation in which the diamond-shaped object was absent.

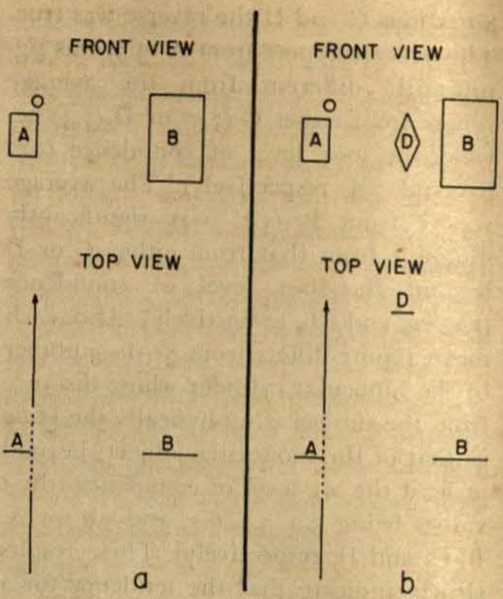


FIG. 4. Schematic front and top-view drawings of the situations used to investigate the effect of the presence of a diamond-shaped figure (D) upon the apparent depth between two differently sized playing cards (A and B).

Procedure

As in the first two experiments of this study, the stereoscopically generated disc was used to measure apparent depth, in this case, between the two playing cards. As indicated in Fig. 4, the line-of-sight position of the disc was always such that the line of sight to the disc was above (10′ of visual angle) the line of sight to the normal-sized playing card. Eight subjects were used. The subjects were asked to adjust the disc to apparent depth equality with each of the playing cards, both when the diamond-shaped object was present and when it was absent. The bracketing technique was used, with the final distance adjustment of the disc systematically varied with respect to whether the disc moved toward or away from the subject. The procedure was systematically varied between subjects with respect to (*a*) whether the situation represented by Part *a* or by Part *b* of Fig. 4 was presented first, and (*b*) whether the left or right card was designated first as the object to which the disc was to be adjusted in distance. No objects were visible except the disc, the two playing cards, and the diamond-shaped figure when it was present.

Results

The summarized results from this experiment are shown in Table 4. Each mean in the table is an average of eight scores, one from each subject, where each score is the mean of 16 adjustments of the disc to a particular playing card. The standard deviations of Table 4 were computed from the distributions of eight scores.

It will be seen from the mean results of Table 4 that when the diamond-shaped figure was absent, the average apparent depth between the two playing cards, as measured by using the disc, was 23 cm. (305 cm. minus 282 cm.), with the double-sized card appearing in *front* of the normal-sized card. But when the diamond was present, the average apparent depth between the two playing cards, as measured by using the disc, was 8 cm. (312 cm. minus 304 cm.), with the double-sized card slightly in *back* of the normal-sized card. The difference (31 cm.) between these two pairs of results is significant beyond the .01 level of confidence ($t = 3.9$). The average verbal report was that the right card was 11 in. in front of the left card when the diamond was absent, and 1 in. in front of the left card when the diamond-shaped figure was present. It seems that the inverse relation between lateral separation and the strength of the tendency to see a monocular object at the apparent distance of a binocular object is capable of significantly modifying the apparent depth difference produced by relatively strong size cues.

TABLE 4

MEANS AND STANDARD DEVIATIONS IN CENTIMETERS OF THE ADJUSTMENT OF A BINOCULAR DISC TO APPARENT EQUIDISTANCE WITH TWO DIFFERENTLY SIZED PLAYING CARDS DEPENDING UPON WHETHER A DIAMOND-SHAPED FIGURE (D) IS PRESENT OR ABSENT

Distance adjustment of binocular disc with respect to:	D Absent		D Present	
	Mean	*SD*	Mean	*SD*
Normal Card	305	6.4	304	5.7
Large Card	282	17.2	312	30.3

Discussion

It appears that the increase in the effectiveness of the size cue to relative depth with increased lateral separation found in the previous study (3), when one of the playing cards was monocularly and the other binocularly viewed, can be explained by the inverse relation be-

Bureau Ednl. Psy. Research
DAVID HA[illegible]AINING COLLEGE
Dated
Accs No

tween lateral separation and the strength of the tendency to see objects as equidistant. Increasing the lateral separation of the playing cards reduced the restrictive influence of the tendency to see the playing cards as equidistant and, as a result, the size cue between the cards had an increasingly important role in determining the relative distance perception. Similarly, it also might be expected that the effectiveness of other monocular cues of relative distance, such as brightness differences between objects, would be a function of the lateral separation of these objects.

GENERAL DISCUSSION

In a situation in which no cues to the relative depth location of objects are present, the objects probably will tend to be seen as equidistant, regardless of their lateral positions. This is expected to occur, even though the evidence is that the strength of the tendency to see objects as equidistant decreases as the lateral line-of-sight separation of the objects is increased. If there are no opposing factors, even the weak equidistance tendencies resulting from large lateral separations will probably be effective.

In the second experiment the binocular disparity cue was present to produce the apparent depth between the two squares. The monocular rectangle had no strong relative distance cues to determine its depth position with respect to the squares. To the extent that the monocular rectangle appeared to be closer in depth to the laterally closer square, the tendency for the rectangle to appear at the depth of the laterally closer square was stronger than the tendency for it to appear at the depth of the other square. In the third experiment, binocular disparity cues were present to produce the apparent depth between the binocular cylinders. The size cue resulting from similarity of shape was present between the monocular cylinder and the binocular cylinders. In spite of this, the monocular cylinder appeared to be displaced from its physical relative depth position toward the depth position of the laterally closer binocular cylinders. In the fourth experiment, binocular disparity cues were present to produce apparent depth between the diamond-shaped figure and the left (binocular) playing card. Despite fairly strong size cues between the binocular and monocular playing cards, the relative depth position of the monocular playing card differed, depending upon whether or not the diamond-shaped figure was present. These last three experiments of the study represent the type of situation in which the apparent distance aspects of parts of the field of view are fairly well determined. It is found that the apparent depth position of an additional object with less strong distance cues can be influenced by the inverse relation between lateral line-of-sight separation and the tendency to see objects as equidistant.

SUMMARY AND CONCLUSIONS

It was demonstrated quantitatively that there is a tendency for a monocular object to be seen at the same apparent distance as a binocular object. More generally stated, the conclusion would be that two objects tend to be seen at the same apparent distance. The presence of strong binocular cues, however, can severely restrict or eliminate the effectiveness of this factor. Experimental results indicate that the strength of the tendency for objects to appear equidistant decreases as the lateral line-of-sight separation of the objects is increased. This was demonstrated in several visual situations, and the tendency seems to be sufficiently strong to modify relative distance perceptions in situations involving size cues. These results were used to "explain" the increase in the effectiveness of size cues with increased lateral separation, which was found in that part of a previous study (3) in which stereopsis was effectively absent between the two similar but differently sized objects.

REFERENCES

1. ALLUISI, E. A., & HARKER, G. S. The stereoptometer—a simple haploscopic instrument for the study of binocular space perception. *Science,* 1953, **117**, 682-683.
2. GOGEL, W. C. Perception of the relative distance position of objects as a function of other objects in the field. *J. exp. Psychol.,* 1954, **47**, 335-342.
3. GOGEL, W. C., & HARKER, G. S. The effectiveness of size cues to relative distance as a function of lateral visual separation. *J. exp. Psychol.,* 1955, **50**, 309-315.
4. JUDD, C. M. An optical illusion. *Psychol. Rev.,* 1898, **5**, 286-294.
5. TSCHERMAK-SEYSENEGG, A. *Introduction to physiological optics.* (Trans. by P. Boeder.) Springfield, Ill.: Charles C Thomas, 1952.

(Accepted for publication September 19, 1955)

Vol. 70, No. 5 Whole No. 412, 1956

Psychological Monographs: General and Applied

Absolute and Relational Stimulus Training in Discrimination Learning[1,2]

Sheldon J. Lachman[3]
Wayne University

I. Introduction

A. *Viewpoints Regarding the Nature of Discrimination Learning*

The differential reactions of animals to stimuli differing quantitatively from each other along a particular dimension have long and often been regarded as a function of the relational nature of the stimulus situation. Among the early investigators in the field, Kinnaman in 1902 (8), Casteel in 1911 (2), Bingham in 1913 (1), and Coburn in 1914 (3), suggested that in discrimination problems, the animal may learn to respond to the relational characteristics of the stimulus situation rather than to the specific properties of each stimulus object. Although these researchers worked and published their statements before the development of gestalt psychology, this notion is generally regarded as a contribution of that school. Köhler is frequently credited with being the modern proponent of the concept, which is traced back to an article on brightness discrimination in chicks that was published in 1918 (9). More recently Lashley (12), Gulliksen and Wolfle (4, 5), Krechevsky (10), and Lashley and Wade (13) have advocated this concept.

Several other psychologists have been critical of the relational interpretation. Among these are Gundlach and Herington (6), Taylor (18), Warden and Rowley (19), and Warden and Winslow (20). But very few of the critics have proposed constructive alternative interpretations. An exception has been Spence, who has developed a theoretical scheme based on stimulus-response principles and concepts (16, 17). Spence's theory and that of Hull (7) are in many respects very similar.

According to Spence's hypothesis, in his own words:

> . . . discrimination learning is conceived as a cumulative process of building up the strength of the excitatory tendency of the positive stimulus cue (*i.e.*, the tendency of this stimulus to evoke the response of approaching it) by means of the successive reinforcements of the responses to it, as compared with the excitatory strength of the negative stimulus, responses to which receive no reinforcements (17, p. 431).

---

[1] This monograph is based on a dissertation submitted to the University of Michigan in partial fulfillment of the requirements for the degree of Doctor of Philosophy, August, 1951.

[2] A summary of this investigation was presented at the 1953 convention of the Midwestern Psychological Association in Chicago.

[3] The writer wishes to express his thanks to Professors Carl R. Brown, Lee R. Dice, Donald G. Marquis, Burton D. Thuma, and Dr. Donald W. Lauer, members of the doctoral committee, for their helpful criticisms and suggestions, and especially to Professor John F. Shepard, the chairman of the committee, who gave generously of his time, suggestions, criticisms, and encouragement throughout the planning and accomplishment of the research. The author also wishes to thank Dr. Herbert S. Conrad and Dr. John F. Dashiell whose constructive criticisms in the preparation of the present manuscript have resulted in its improvement.

Bureau Ednl. Psy. Research
[illegible] COLLEGE
Dated

B. *A Few Comments on the Problem of Stimulus Control*

A primary objection to certain research investigations of discrimination learning relating to the relational and absolute stimulus hypotheses is that the stimuli involved varied in more than a single dimension. Each stimulus actually provided opportunity for responding to two or more aspects or relationships in the situation. This lack of control is an inherent characteristic of a large number of complex stimuli. There is little doubt that stimuli which are distinctive from each other and which bear several differing relationships to their fields can be responded to in absolute terms.

To refer specifically to the matter of visual size as a stimulus: Several problems are presented if one attempts to vary the size of even a simple white circle on a black card of constant size. For example, if its size is to be doubled, does this mean that the stimulus (white) area is to be doubled? If the area is so increased, then is the size of the stimulus really doubled? The horizontal diameter has not been doubled, nor has the vertical diameter. If the horizontal diameter is doubled, even without doubling the vertical, then the stimulus is no longer a circle but an ellipse. If both diameters are doubled, then, of course, the stimulus is much more than doubled in total area.

Let us consider the matter more generally with an illustration: Frequently when size is utilized as the stimulus "dimension" to be discriminated, in the selection of the second stimulus more than one aspect of the stimulus situation is simultaneously changed, i.e., (*a*) the total *area* of the stimulus is modified, (*b*) the total quantity of illumination is altered, (*c*) the distance between the floor and base of the stimulus object or between the top of the stimulus object and the ceiling is (or both are) changed, (*d*) the distances between the stimulus-object and the peripheries of the field are altered, (*e*) there is a change in the size of the stimulus object (increase or decrease) relative to the size of the visual field. The animal may be responding to one or more of these variations or relationships. *Something of a patterned character exists in each stimulus.*

C. *Purpose of this Investigation*

The purpose of this investigation was to obtain evidence pertaining to the adequacy of relational and absolute stimulus-training techniques for discrimination learning by employing a method which involved a greater degree of stimulus control than has usually been achieved in research on discrimination. At the same time, relative effectiveness of learning under these two sets of conditions would provide evidence relevant to the theoretical positions on discrimination learning proposed by Köhler and by Spence.

II. THE DESIGN OF THE INVESTIGATION

A. *The Stimuli*

For the proposed investigation it was necessary to select simple stimuli which varied only in degree from other stimuli of the same class. To put the matter another way, the two stimuli to be employed were each to be as elementary as possible and the difference between the positive and the negative stimulus was to be confined to variation in degree of a single, not a multiplicity, of aspects. Such stimuli when presented singly, that is to say, in isolation, would then be simple and patternless, i.e., there would be no relationships apparent within a single stimulus. When presented together, however, these two stimuli would form a pattern, for according to gestalt psychologists, not only are the properties of each individual stimulus present, but also an additional quality or property dependent upon relational or comparative aspects of the two simultaneously present stimuli, in terms of which the animals may learn.

Stimuli which vary in brightness only are relatively simple, relatively convenient to employ, and are relatively easily controlled. Since such stimuli when employed singly, in isolation, fulfilled the requirements of the associationistic-type theory which proposes that discriminative learning is a function of the absolute properties of each stimulus, and since two such stimuli in combination fulfill the criterion of *patterned,* i.e., relational, stimuli for the gestalt-type theory, these have been utilized as the stimuli in this investigation.

B. *The Subjects*

1. *Number and ancestry.* The subjects (Ss) were 42 mature female albino rats obtained from the laboratory of Dr. E. L. Walker, at the University of Michigan. Group I consisted of eight animals, Group II of ten, Group III of 12, and Group IV of 12. Dr. Walker's rat colony originated with animals purchased from three commercial animal farms. Regarding the ancestry of the Ss, little is known except that attempts at selective breeding were confined to mating healthy animals, i.e., animals that gave no superficial signs of illness.

2. *Age and naïveté.* All Ss were naïve in the sense that none had been involved in any previous experimentation, psychological or otherwise. All animals were two to six months of age at the beginning of the experiment.

3. *General living conditions during the experiment.* The Ss were kept in a room separate from, but adjacent to, the experimental room.

After completing their trials for any day, animals were put into a neutral cage for 15 minutes and then transferred to a food cage where they were permitted to eat for about 30 minutes. The diet consisted of Purina Laboratory Animal Chow, supplemented weekly by green vegetables.

Animals in Groups I and II were employed successively, that is to say, work with Group II animals was not initiated until work with Group I animals was almost completed. The members of these two groups, while matched for weight, sex, age, and equivalence of background experience, were not, of course, from the same litters.

Animals in Groups III and IV were utilized at the same time, i.e., run on the same days, and these groups consisted of matched members of split litters. These animals were matched for weight, sex, age, and equivalence of background experience.

C. *Design of the Apparatus*

The apparatus was designed to permit variation in one stimulus aspect, viz., brightness, while maintaining others invariable and constant (11). The instrument employed was a modified Yerkes-type brightness discrimination apparatus.[4] It can be best described with the aid of Fig. 1, a floor plan diagram.[5]

1. *The discrimination box.* A tunnel, A, led to the discrimination box proper, B. A low partition one inch high, E, functioned to impede

---

[4] A more complete description of the apparatus and procedure (including additional diagrams), together with data on the performance of individual subjects, has been deposited with the American Documentation Institute. Order Document No. 4758 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D.C., remitting in advance $2.00 for microfilm or $3.75 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

[5] Capital letters used from this point through the section to indicate parts of the apparatus refer to Fig. 1, Floor-Plan Diagram of the Apparatus.

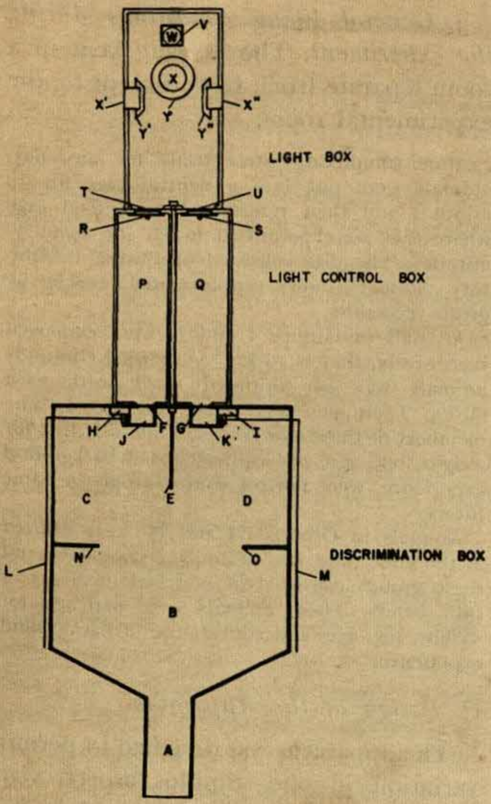


Fig. 1. Floor-plan diagram of brightness discrimination apparatus.

the attempts of the animal to move from one stimulus compartment to the other, C and D, but permitted the animal readily to observe the stimulus in the adjacent compartment. Stimulus windows, F and G, were square and of flashed opal glass. At the lateral aspect of each stimulus window was a food release mechanism, H and I, which deposited a single pellet at a time into a food trough, J and K, directly beneath the stimulus window when operated mechanically via two metal rods, L and M, by the investigator from a position near the tunnel. Two narrow vertical strips of metal, one on each side of each stimulus window, provided narrow grooves into which a thin metal plate, the stimulus window obstruction plate, could be fitted tightly so as to obscure completely the stimulus window and prevent any light from passing through it into the stimulus compartment. (These and many other details of the apparatus are diagrammatically represented in Fig. 2.)

2. *The light-control box.* The light-control box was located directly behind the discrimination box. This box enclosed two chambers, P and Q, separated from each other by a light-tight partition; the interior of each of these chambers was painted black. In the posterior wall of each chamber was a pair of shutter plates, R and S. The shutters were made of metal and painted black. A brass rod 1/8 of an inch in diameter passed vertically through the two metal blocks, one attached to the lower and the other to the upper shutter of each system and projected outside the light box. Each brass rod was so threaded that by appropriately turning it, the plates of the shutter system could be made to separate from or approach each other. This rod was called the light-regulation screw. Thin and separate sheets of uniformly bilaterally sanded glass, T and U, rested against each pair of shutter plates.

3. *The light box.* The most posterior structure was the light box. This was lined with sheet aluminum and contained a single 500-watt projection lamp, W, located on a small metal base, V. A ventilation system for the light box consisted of circular orifices in the floor, ceiling, and wall, X, X′ and X″, which were baffled by blackened sheets of metal, Y, Y′ and Y″. This arrangement permitted heat to escape but prevented light from doing so, and permitted the maintenance of a closed lighting system.

4. *The voltage stabilizer.* A voltage stabilizer was employed to guard against momentary voltage fluctuations and to maintain the voltage at 120 volts at the light box.

5. *Surface of the apparatus.* Except for a small part of the choice chamber, which was painted a medium gray, and the interior of the light box which was lined with aluminum, all of the apparatus, both inside and out, was painted flat black.

6. *Table.* The apparatus rested upon and was fastened by means of screws to a blackened heavy wooden table.

7. *The canopy.* The apparatus and table were enclosed within a canopy consisting of a wooden framework completely covered by black cloth.

8. *The lighting system.* The source of illumination for both stimulus windows was a single 500-watt lamp in the light box. Light emanating from this source passed through a sheet of uniformly sanded glass, through the diaphragm of the light-control box, then diffused through the two foot-long blackened corridors of the light box, and illuminated each flashed opal stimulus window with a high degree of homogeneity. A beveled type of window environment was used to ensure uniform illumination of each window and to obviate the possibility of shadow at the borders of the stimulus windows (see Detail A, Window Detail, of Fig. 2). Without this precaution, depth of shadow at the periphery of the window would vary with

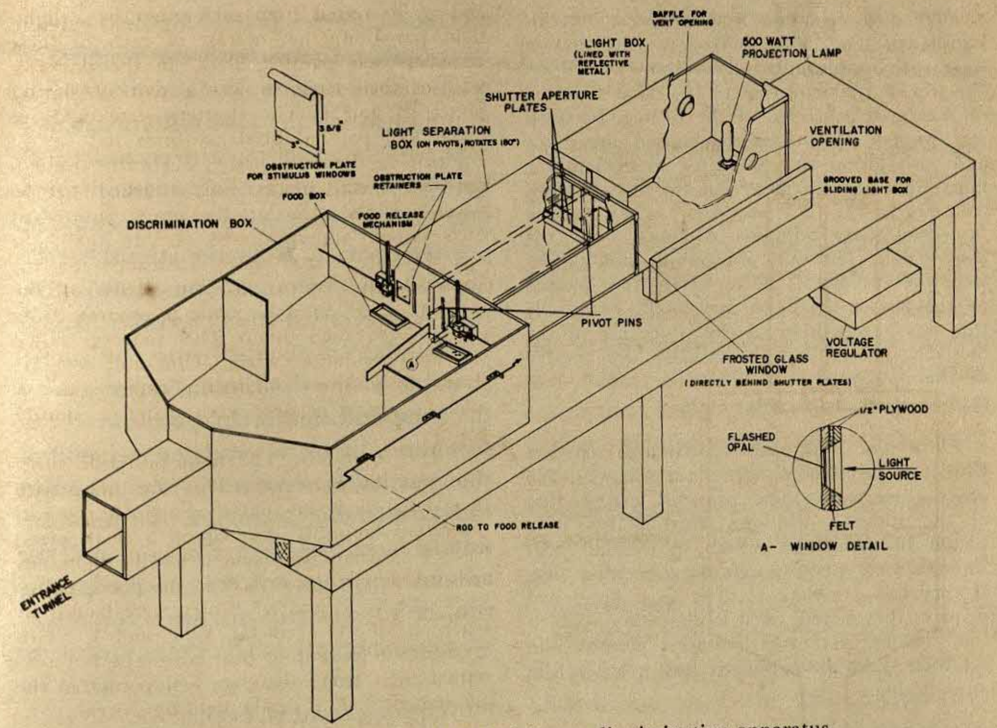


FIG. 2. Isometric diagram of the brightness discrimination apparatus.

the intensity of illumination. Contrast effects between shadow and general illumination of the stimulus might also occur. These could serve as unintended cues to the *Ss*. Furthermore, these uncontrolled variables might provide a possibility for learning to respond to individual stimuli by relational rather than absolute means.

9. *Measuring the brightness of the stimulus windows.* By means of a Weston Photoronic Cell and a microammeter and allied apparatus, the physical brightness of the stimulus windows for various turns of the light-regulation screw were determined in terms of foot-lamberts. From zero to 18 revolutions of the light-regulation screw an approximately rectilinear representation was obtained. In this paper it will be more convenient to indicate brightness by specifying the number of revolutions of the light-regulation screw which are integers than by specifying the brightness in terms of foot-lamberts, which demands the use of small fractions of units.

D. *Motivation*

Reinforcement was provided by food pellets, disclike tablets, composed of Purina Laboratory Animal Chow, each of which was slightly smaller than a standard size aspirin and each of which weighed approximately one centigram. Animals were run daily with an interval of 22 to 26 hours elapsing between each day's complement of ten trials.

E. *Procedure*

1. *Pre-experimentation handling.* The groups of animals were obtained about a month prior to the beginning of training of each group and had not been handled up to that time. With constant daily handling, defecatory, micturitional, and other diffusive behaviors progressively diminished.

2. *Preliminary training.* Groups I and II received no preliminary training. In an effort to reduce the time and trials required for reaching the criterion of learning, Groups III and IV were given four days of preliminary training. This preliminary procedure involved the following:

a. First day. Stimulus windows were not illuminated. The only sources of illumination were two 8-watt bulbs which were located on the top of the light-control box, and the rays of these were directed at the ceiling. No food was present in the discrimination apparatus and each group was permitted to explore the apparatus, a group at a time, for a half hour.

b. Second day. Stimulus windows were not illuminated. The only sources of illumination were the two 8-watt bulbs which were located on the top of the light-control box. Essentially this was a repetition of the procedure of the first day, except that the animals were placed in the apparatus individually and allowed to explore it for 15 minutes each.

c. Third day. Both stimulus windows were set with the positive stimulus, the dimmer. No other illumination was provided except that of the stimulus windows. Food pellets were placed in each food trough in a small pile. Animals were permitted to eat on either side. If they had not done so after 10 minutes had elapsed, they were guided to the side on which they had not yet eaten and were permitted to eat there. Each animal ate at least a few pellets on each side.

d. Fourth day. Both stimulus windows were set with the positive stimulus. No food was in the food troughs. Animals were put in at the tunnel and were given ten runs to the stimulus windows. The first five runs were unguided. But as many of the last five runs were guided as necessary to achieve a symmetry of five runs to the right and five runs to the left stimulus compartment. A single pellet was released after each run and the animal was permitted to eat it.

3. *The training of Group I.* The stimuli to be discriminated were the stimulus windows of flashed opal illuminated by an arbitrary setting of the light regulation screw, at one and seven complete revolutions. In terms of the physical measures used, these two stimuli possessed a brightness *ratio* also in terms of foot-lamberts of one to seven. The stimulus windows provided the only source of illumination for the animals in the situation. No other light source was permitted, and since the background was dark due to the unlighted room, the shrouding canopy, and the black surface of the apparatus, little light was reflected.

a. General procedure. For this group, the positive stimulus was the brighter; a relational technique was used, i.e., a technique in which the possibility existed of comparing or relating the stimuli since both were simultaneously present. On a given trial the rat was placed backwards, i.e., facing away from the stimulus windows in the entrance tunnel. Running time was recorded from the time all four feet of the rat emerged into the choice chamber of the discrimination box until the head and forefeet passed toward the stimulus window beyond the free end of the partition separating the stimulus compartments. A run this far was considered a choice or a decision and the animal was not permitted to retrace his run, or to cross the partition, or otherwise to emerge from the stimulus compartment until he was manually removed for the next trial.

If the animal's choice was to the positive stimulus compartment, a food pellet dropped from the food-release mechanism into the food trough directly in front of the stimulus window. A rod from the food release extended to the outside of the discrimination box and connected to another rod which ran laterally toward the "front" part of the discrimination box (behind the rat) where the experimenter stood (see Fig. 2). A slight movement of the experimenter's finger was sufficient to release mechanically a single pellet into the food trough. This device functioned in a highly efficient manner and there was little latency in the system. The animal was required to remain in the stimulus compartment, facing the stimulus window, for 10 seconds. If the animal's choice was to the negative stimulus compartment, the animal received no food. However, the animal was required to remain in the compartment for 10 seconds.

About 15 to 25 seconds elapsed between the completion of one run and the beginning of the next. Decision time and correctness or incorrectness of response were recorded during this period. Ten trials with the stimuli varying in a randomized fashion from trial to trial constituted the daily training. Animals were run until they reached the criterion of training, and for a varying number of days thereafter.

b. The criterion of learning. The criterion of training for this experiment was perfect performance on four successive days. In other words, in order to reach the criterion the animal was required to perform at a 100 per cent level of success, making no errors for four successive days with 10 trials per day. This was regarded as an adequate and rigorous criterion.

(i) Evaluation of the learning criterion according to the probability theory. In terms of probability concepts, the chance of this level of success occurring is extremely remote, 1 chance in 2^{40}.

(ii) Evaluation of the learning criterion according to currently employed standards. In terms of meeting the standards set to date by comparative psychologists, this criterion appears to be adequate. Munn, in discussing learning criteria of discrimination experiments (15), makes this summary type statement: "When the animal discriminates between the two stimuli with an accuracy of anywhere from 80 per cent to 100 per cent for from 30 to 50 successive trials, the problem is usually regarded as learned." The criterion, as indicated above, falls within the ranges specified by Munn.

(iii) Evaluation of the criterion by empirical test. An empirical test of the criterion was likewise attempted. Animals who had reached the criterion were run for additional days in order to obtain evidence relevant to the adequacy of the criterion, i.e., in order to be more certain that the performance of the animals was a relatively stable modification of behavior.

4. *The training of Group II.* In training, the same stimuli were used for Group II as for Group I, and again the positive stimulus was the brighter. Training, however, was by the absolute technique which did not permit a simultaneous comparison of the aspects of the component stimuli. Only a single stimulus, either the positive or negative, was exposed at a time. Since the animal could view but one stimulus at a time, there was no opportunity for direct comparison of the stimuli used. The stimulus exposed was sometimes on the left and sometimes on the right, the order being randomized. Exposing but a single stimulus was accomplished by sliding a black obstruction plate, the stimulus window obstruction plate, in front of one stimulus window. This served to obscure completely that window and to create a homogeneous black field in that stimulus compartment. On each day's complement of trials, the negative stimulus was exposed three times, the positive stimulus seven times. If the animal ran to the stimulus compartment in which the negative stimulus was exposed, the animal received no reinforcement. It was delayed in the compartment for about 10 seconds. It was then put on a waiting platform for about 15 to 25 seconds, during which time the stimulus windows were appropriately altered for the next trial. If the animal responded positively, it was reinforced with a single food pellet which dropped into the trough directly in front of the stimulus window. Runs to the nonstimulus compartment were, of course, nonreinforced, although they were recorded.

The same standards and procedures in terms of length of time permitted in the tunnel, length of time permitted in the choice chamber, etc., applied to Group II as applied to Group I.

The animals in Group II were given 10 trials daily as indicated above. In terms of total trials, each animal was run until it had received at least as many reinforcements with the positive stimulus and as many reinforcements with the negative stimulus as were required for the poorest, i.e., slowest animal in Group I to reach the criterion of learning. In actual practice each animal made 600 runs. According to the Spence concept of discrimination learning, such training with the individual stimuli would provide for the development of strong positive reactions (approach tendencies) to the reinforced stimulus and of strong inhibition of approach reactions with regard to the nonreinforced stimulus. This theory would hold that discrimination is essentially a matter of building up specific habits of response to each component of the paired stimulus situation. Hence, ability of the animals in this group to make discriminations appropriately in the *test* trials subsequent to training would tend to support the associationistic or absolute stimulus hypothesis; failure to do so, however, would lend support to the gestalt-type or relational stimulus hypothesis.

5. *The training of Group III.* Twelve albino female rats constituted this group. The stimuli to be discriminated for Groups III and IV were the same as the stimuli provided for Groups I and II, i.e., stimulus windows of flashed opal were illuminated by arbitrary settings of the light regulation screw at one and seven complete revolutions.

For Group III the positive stimulus was the dimmer; a relational training technique was employed. Except for the fact that these animals received four days of preliminary training, and were trained to the dimmer rather than to the brighter stimulus, they were handled and treated exactly like the animals in Group I. The same criterion of learning, length of time permitted in the tunnel, length of time permitted in the choice chamber, etc., applied to group III as to Group I.

6. *The training of Group IV.* Twelve female albino rats constituted this group. For Group IV the positive stimulus was the dimmer; an absolute training technique was used (previously described when Group II training was considered). Except for the fact that these animals received four days of preliminary training, they were handled and treated in exactly the same way as the animals in Group II.

The animals in Group IV were given ten trials daily as indicated above. Each animal was run until it had received at least as many reinforcements with the positive stimulus and as many nonreinforcements with the negative stimulus as were required for the poorest (i.e., slowest) *S* in Group III to reach the criterion of learning. In actual fact, each animal made 600 runs. According to the Spence concept of discrimination learning, such training with the individual stimulus would provide for the development of strong positive reactions (approach tendencies) to the reinforced stimulus and for the development of strong inhibition of approach reactions with regard to the nonreinforced stimulus. It would hold that discrimination is essentially a matter of building up specific habits of response to each component of the paired stimulus situation. Hence, ability of the animals in this group to make such discriminations appropriately in the *test* trials subsequent to training would tend to support the

Spence or absolute stimulus hypothesis; failure to do so, however, would tend to support the gestalt-type or relational stimulus hypothesis.

F. *Special Precautionary Techniques Instituted for this Research*

In order to minimize the possibility of unintentional and extraneous stimuli serving as guidance cues to the animals, the following precautionary techniques were instituted:

1. *Attire of experimenter: minimizing visual cues from the experimenter.* During each trial the experimenter (*E*) stood motionless at one particular place, behind the discrimination box (i.e., opposite to the wall in which the discriminable stimuli were situated). The *E* wore a full-length laboratory coat which was dyed black, a black mask over his face which contained small slits at eye level to permit vision, and tight-fitting skullcap. The purpose of this attire was to minimize the possibility of *E* communicating cues, however unintentional, to the *S*s. It was believed that little if any light was reflected from *E* under these conditions and, therefore, movements, even unintentional guide cues, were unlikely to be communicated to the animals.

2. *Disorientation of animals at the beginning of each trial: minimizing handling cues from E.* At the beginning of each trial the animal was placed backwards (i.e., with its head toward the entrance) in the tunnel. This provided for a temporary disorientation in space which facilitated the dissipation of any tactual-kinesthetic cues that might have been given inadvertently in handling the animals; it is *E*'s belief that there was almost no variation in the handling of the animals from trial to trial.

3. *Homogeneity of the environment: reduction of cues from the apparatus and surroundings.* The apparatus rested upon and was fastened by means of screws to a wooden table previously indicated. The apparatus and table were enclosed within an artificial darkroom, a canopy of black. This was constructed of a wooden framework painted black. Fastened to the framework by thumb tacks on all sides including the ceiling (but not the floor) was black cloth.

The experimental room in which the canopy and other equipment were placed was completely darkened. Since the only source of light in this situation was that of the stimulus windows, it is believed that there was no opportunity for responding to other stimuli than these, at least other visual cues.

4. *Randomization in the presentation of the stimuli.* In order to minimize the possibility of a position habit being rewarded more than a chance number of times, a random order of presentation of the stimuli was utilized. The following 16 different relatively random series of stimulus presentations were employed on successive days. These were as follows: LRRLRRLRLL, RLRLLRRRLL, LLRRLRLLRR, RLLLRRLRLR, LRLLRRLRRL, LRLRLLRRLR, RLLRRLRRLL, RRLRLLLRLR, LRLRRLLLRR, RRLLRLRRLL, RLLRLLRLRR, LRRRLLRLRL, RLRLRRLLRL, LRRLLRLLRR, RLRRLLRLLR, LLRLRRRLRL.

In the case of Groups I and III the R or L indicated the position of the positive stimulus. Groups II and IV, however, were given absolute stimuli, i.e., were trained with only a single stimulus exposed, sometimes the positive and sometimes the negative. The same series was used for Groups II and IV, but, of every 10 stimuli exposed in each day's complement, the positive was exposed several times, the negative three times for Groups II and IV. The sequences were as follows: A minus (—) sign follows the designation of position to indicate that the exposed stimulus was negative; all others are positive. LR — RLR — RL — RLL, RLRL — L — R — RRLL, LLR — R — LRL — LRR, RL — LLR — RLRL —R, L — RL — LRRLRR —L, L — RLRLLRR — LR —, RL — LRR — LRRLL—, RRLR — LL — LR — LR, LRL — R — RLL — LRR, RR — L — LRLRR — LL, RLLRL — L —RLRR —, LRR — RLLR — LRL—, RL — RL — RR — LLRL, LRR — LL — RLLRR —, RLR — RLL —RLL — R, LLRLRRR — LR — L—.

5. *Reduction of food cues.* Food appeared to be comparatively odorless. There were an approximately equal number of food pellets in the tubes of the food-release mechanisms of either stimulus compartment at all times. Furthermore, a differential quantity of food on either side, even if detectable olfactorily, would provide no differential cue regarding the correct stimulus window since food was unavailable to the animal until the experimenter released it, and food was not released until *after* a correct choice had been made.

III. THE RESULTS OF THE INVESTIGATION

A. *The Primary Results of the Investigation*

The main results of this investigation are contained in Tables 1, 2, 3, 4, 5, and 6; certain closely related supplementary data which summarize the accuracy of postcriterion trials are contained in Tables 7 and 8.

The primary results are indicated below:

1. *Group I results.* The *S*s in Group I learned to respond positively to the brighter of two stimulus windows under conditions of relational stimulus training. *All* ($N = 8$) *subjects attained the criterion of learning,* which consisted of perfect runs of four successive days of 10 trials per day (40 consecutive errorless responses) in from 210 to 390 trials. The mean number of trials (both reinforced and nonreinforced) required for reaching the criterion was 316.3 (see Table 1).

The slowest learner in Group I, in terms of the number of reinforced responses to the positive stimulus prior to learning, required 279 reinforced trials; the slowest learner in Group I, in terms of the number of nonreinforced responses to the negative stimulus prior to learning required 132 nonreinforced trials.

2. *Group II results.* The *S*s in Group II were trained under conditions that did not permit relational learning, i.e., by the absolute stimulus method that did not permit comparison between the stimuli, and each of these was given more positive (reinforced) and negative (nonreinforced) trials than the slowest learner in Group I. Hence, in comparison with Group I results, each animal in Group II received 346 or more reinforced trials to the positive stimulus; each animal in Group II received 144 or more nonreinforced trials to the negative stimulus (see and compare Tables 1 and 2).

TABLE 1

LEARNING DATA FOR GROUP I SUBJECTS WHO WERE TRAINED BY THE RELATIONAL METHOD WITH THE BRIGHTER STIMULUS AS POSITIVE

	Trials Required for Learning	No. of Responses to Positive Stimulus (Reinforced Trials)	No. of Responses to Negative Stimulus (Nonreinforced Trials)
Range	210–390	149–279	61–132
Mean	316.3	205.6	110.6
SD	59.8	43.4	21.9

TABLE 2

LEARNING DATA FOR GROUP II SUBJECTS WHO WERE TRAINED BY THE ABSOLUTE METHOD WITH THE BRIGHTER STIMULUS AS POSITIVE

	No. of Responses to Positive Stimulus (Reinforced Trials)	No. of Responses to Negative Stimulus (Nonreinforced Trials)	No. of Responses to Nonstimulus Compartment
Range	346–408	144–176	18–110
Mean	377.2	161.5	61.3
SD	18.6	9.4	27.7

Most of the *S*s in Group II ($N = 10$) reached or closely approximated the criterion of learning (see Table 3). Three animals in this group, however, made 4, 11, and 13 errors on their respective test trials (see Table 3).

3. *Group III results.* The *S*s in Group III learned to respond positively to the dimmer of two stimulus windows under conditions of relational stimulus training. All ($N = 12$) *S*s reached the criterion of learning which consisted of perfect runs on four successive days of 10 trials per day in from 240 to 430 trials. The mean number of trials required for learning was (both reinforced and nonreinforced) 309.2 (See Table 4).

The slowest learner in Group III, in terms of the number of reinforced responses to the positive stimulus prior to learning, required 269 reinforced trials; the slowest learner in Group III, in terms of the number of nonreinforced re-

TABLE 3

ERRORS OF GROUP II SUBJECTS ON TEST TRIALS WITH BRIGHTER STIMULUS AS POSITIVE

	Test Trials				Total No. of Errors	Per Cent of Success
	1st Day (Trials 1–10)	2nd Day (Trials 11–20)	3rd Day (Trials 21–30)	4th Day (Trials 31–40)		
Total	6	10	8	6	30	
Range	0–5	0–5	0–4	0–3	0–13	67.5–100
Mean	0.6	1.0	0.8	0.6	3.0	96.3

sponses to the negative stimulus prior to learning, required 161 nonreinforced trials.

4. *Group IV results.* The *Ss* in Group IV were trained under conditions that did not permit relational learning, i.e., by the absolute method, which did not permit comparison between the stimuli, and each was given more positive (reinforced) trials and more negative (nonreinforced) trials than the slowest learner in Group III. Hence, in comparison with Group III results, each animal in Group IV received 327 or more reinforced trials to the positive stimulus; each animal in Group IV received 163 or more nonreinforced trials to the negative stimulus, except one subject of this group who through an oversight received only 159 nonreinforced trials (see Tables 4 and 5).

The subjects in Group IV ($N = 12$) did not even remotely approximate the learning criterion (see Table 6). With the exception of one animal that obtained a score of 50 per cent success on the test trials by adopting a position habit, no *S* exhibited more than 45 per cent success on the test trials. And one animal was successful only 2.5 per cent of the time, which represents one correct run of the 40 test trials. Hence, the Group IV animals, like the animals in Group II on being tested, displayed from slight to very strong preferences for the brighter stimulus, although these animals had received absolute training to the dimmer, and had been reinforced during the training only with the dimmer stimulus.

One *S* of this group displayed a position habit from the very first test trial. Preferences for the negative, i.e., the brighter, stimulus existed in all other *Ss* of this group on the initial trials, after which rigid stereotyped position re-

TABLE 4

LEARNING DATA FOR GROUP III SUBJECTS WHO WERE TRAINED BY THE RELATIONAL METHOD WITH THE DIMMER STIMULUS AS POSITIVE

	Trials Required for Learning	No. of Responses to Positive Stimulus (Reinforced Trials)	No. of Responses to Negative Stimulus (Nonreinforced Trials)
Range	240–430	140–269	87–161
Mean	309.2	191.5	117.7
SD	47.0	45.5	18.8

TABLE 5

LEARNING DATA FOR GROUP IV SUBJECTS WHO WERE TRAINED BY THE ABSOLUTE METHOD WITH THE DIMMER STIMULUS AS POSITIVE

	No. of Responses to Positive Stimulus (Reinforced Trials)	No. of Responses to Negative Stimulus (Nonreinforced Trials)	No. of Responses to Nonstimulus Compartment
Range	372–402	159–172	30–114
Mean	378.8	166.7	54.6
SD	22.3	3.5	25.1

TABLE 6

Errors of Group IV Subjects on Test Trials with DIMMER Stimulus as Positive

	Test Trials				Total No. of Errors	Per Cent of Success
	1st Day (Trials 1–10)	2nd Day (Trials 11–20)	3rd Day (Trials 21–30)	4th Day (Trials 31–40)		
Total	91	77	74	72	314	
Range	5–10	5–9	5–10	4–10	20–39	2.5–50
Mean	7.6	6.4	6.2	6.0	26.2	32.5

sponses were exhibited by several of them.

5. *Empirical test of the learning criterion.* In order to evaluate the adequacy of the criterion of learning employed in this investigation, *S*s in Groups I and III were given a varying number of trials (from 10 to 200) subsequent to their achievement of the criterion. Of the *S*s in Group I, none performed with an accuracy level of less than 92.3 per cent during this postcriterion evaluation period (see Table 7); of the *S*s in Group III, none performed with an accuracy level of less than 90 per cent during this postcriterion evaluation period (see Table 8). These data seem to indicate that the modified behavior is relatively stable and not a transient phenomenon, and as such fulfills an essential requisite of most definitions of learning.

6. *Transposition data.*

a. Transposition training results. Animals in Groups I and III subsequent to their postcriterion trials were trained by the relational method for a period of 15 days, 150 trials, to discriminate between stimulus windows set at 2 and 6 revolutions of the light-regulation screw. For Group I, the brighter, 6, was the positive and, hence, reinforced stimulus. For Group III, the dimmer, 2, was the positive, and hence, reinforced stimulus.

On the first 10 trials of these 150 training trials mentioned above, the stimulus windows were set at 1 and 6 revolutions of the light regulation screws. In Group I, *S*s continued to discriminate with better than 95 per cent accuracy on these 10 trials. From the 11th trial on, the windows were set at 2 and 6. Between the 11th and 20th trials the accuracy level of Group I dropped to 78.8 per cent, but a trend of improvement in the accuracy of response occurred with successive practice during the training period. The accuracy of Group III *S*s fell to 90 per cent on the first 10 trials; it remained at about this level between the 11th and 20th trials, after which a trend of improvement in the accuracy of performance occurred with successive practice during the training period.

Animals then were tested in order to

TABLE 7

Postcriterion Runs for Group I Subjects

	Total No. of Responses	No. of Responses to Positive Stimulus	Per Cent of Accuracy
Range	10–180	10–178	92.5 to 100.0
Mean	82.5	79.3	96.1

TABLE 8

Postcriterion Runs for Group III Subjects

	Total No. of Responses	No. of Responses to Positive Stimulus	Per Cent of Accuracy
Range	10–200	10–183	90.0–100
Mean	130.8	124.3	95.0

determine the extent of transposition by transposition stimuli, the component stimuli of each pair of which bore the same brightness relations to each other as the component stimuli of the standard pair of Stimuli 2 and 6. Hence, animals were tested with one pair of stimuli set at ⅔ and 2 revolutions and another pair set at 6 and 18 revolutions.

b. Transposition test results. The transposition test consisted of 50 trials. Of the 50 trials during the transposition test days, 30 were to the standard stimuli of 2 and 6 upon which the animals were trained. Responses to this pair were reinforced if correct (i.e., if the responses were to the positive stimulus). The purpose of these presentations of the standard pair was to keep the animals trained to the brighter (Group I) or to the dimmer (Group III) stimulus of the pair. Interposed among these standard stimulus presentations were 10 presentations of the stimuli set at ⅔ of a revolution and 2 revolutions, and 10 presentations of the stimuli set at 6 and 18 revolutions. No reinforcements occurred on these test trials regardless of whether the responses were to the positive or negative stimuli.

For each group this meant that on certain trials the originally negative stimulus (negative stimulus of the standard pair) was presented with a new stimulus, and on other trials the originally positive stimulus (positive stimulus of the standard pair) was also presented with a new stimulus. If the animal had been trained to the absolute properties of the stimulus, it should then respond appropriately to the original stimuli by "avoidance" or approach. However, if the *S*s had learned to respond to a relational characteristic (i.e., a comparative aspect) of a pair of stimuli, the *S*s should respond in the opposite way.

Results of the transposition tests are summarized in Tables 9 and 10. It is to be noted that the proportion of errors was approximately the same with the presentations of the new (transposition) pairs of stimuli as it was with the presentations of the training (standard) pairs of stimuli in the test. For both groups the mean success was above 85 per cent (see Tables 9 and 10). These results are interpreted to mean that the animals were responding to a relationship between the stimuli rather than the absolute properties of the individual stimuli.

B. *The Secondary Findings of the Investigation*

Following is a summary of some of the secondary findings of this investigation:

1. *The latency data: reaction times.* Reaction (i.e., decision or choice) time

TABLE 9

ERRORS ON TRANSPOSITION TESTS FOR GROUP I SUBJECTS WHO WERE TRAINED TO RESPOND TO THE BRIGHTER STIMULUS

	Transposition Stimuli		Total No. of Errors (20 trials per subject)	Reactions to Standard Pair
	No. of Revolutions of the Light-Regulation Screw			2 and 6 (30 trials per subject)
	2/3 and 2 (10 trials per subject)	6 and 18 (10 trials per subject)		
Total	6	10	16	27
Range	0–2	0–3	0–4	0–6
Per Cent Accuracy	92.5	87.5	90	88.8

TABLE 10

ERRORS ON TRANSPOSITION TESTS FOR GROUP III SUBJECTS WHO WERE TRAINED TO RESPOND TO THE DIMMER STIMULUS

	Transposition Stimuli: No. of Revolutions of the Light-Regulation Screw: 2/3 and 2 (10 trials per subject)	Transposition Stimuli: No. of Revolutions of the Light-Regulation Screw: 6 and 18 (10 trials per subject)	Total No. of Errors (20 trials per subject)	Reactions to Standard Pair: 2 and 6 (30 trials per subject)
Total	12	24	36	44
Range	0–3	1–3	1–6	0–10
Per Cent Accuracy	90	80	85	87.8

data are of incidental importance in this investigation. At first most *Ss* in all groups made decisions slowly. When the animal hesitated in the choice chamber for more than 30 seconds, it was prodded toward the center and toward the stimulus windows. Such hesitations were most frequent in the early trials. Speed of decision began to increase (i.e., choice time decreased) after varying numbers of trials. On later trials almost all runs were made in one second, arbitrarily the shortest time recorded, or less. A summary of the reaction-time scores for each group is contained in Table 11, which sets forth the data in support of the statements above.

2. *Error scores and learning curves.* The training period for each animal was divided into 10 equal units, each of which, therefore, constituted 10 per cent of the total training period. The number of errors occurring within each of these subperiods was then tabulated, and a distribution of the percentage of errors in each of these subperiods was prepared. The data, grouped in this way (i.e., in terms of Vincent tenths), have permitted plotting of individual learning curves for each animal in Groups I and III in terms of Vincent tenths. The percentage of incorrect responses for each tenth of the training period was indicated. The learning curves derived from results of this investigation are typical of those reported for discrimination learning; they indicate that for the first half to three-quarters of the learning period, error scores vary about (i.e., remain near) the 50 per cent level; then there is a relatively sharp—at least a sharper—downward progression of the curve, indicating a rather sudden increase in accuracy of performance. That is, to say after a long period of no apparent progress, with the results near the chance level, there is a relatively sudden increase in accuracy.

IV. DISCUSSION OF THE RESULTS OF THE INVESTIGATION

A. *Discussion of the Principal Results*

The results of this research tend to support the gestalt hypothesis of discrimination, namely, that discrimination is dependent upon some comparative or relational aspect between the stimulus constituents of a pattern. The results seem to indicate that training with relational stimuli, i.e., with both stimuli of a pair simultaneously present, is not equivalent or reducible to absolute training, i.e., individual training, with only one component stimulus of the pair present at a time. The differences are more

TABLE 11

LATENCY DATA: TABLE OF PERCENTAGES OF REACTION TIME SCORES DURING STAGES OF DISCRIMINATION TRAINING FOR ALL GROUPS

Group No.	Reaction Time Categories (in minutes)	Early Stage (First 30% of Trials)	Intermediate Stage (Middle 40% of Trials)	Late Stage (Last 30% of Trials)
I ($N=7$)*	31 and over	2.6	0.6	0.0
	6 to 30	15.1	2.3	0.1
	2 to 5	37.6	4.7	0.7
	1	44.7	92.3	99.1
	Total	100.0	99.9	99.9
II ($N=10$)	31 and over	13.4	0.0	0.0
	6 to 30	22.6	2.1	0.2
	2 to 5	32.7	8.0	1.4
	1	31.3	89.9	98.4
	Total	100.0	100.0	100.0
III ($N=12$)	31 and over	5.1	0.0	0.0
	6 to 30	15.2	2.2	1.9
	2 to 5	18.9	5.9	6.1
	1	60.8	91.8	92.0
	Total	100.0	99.9	100.0
IV ($N=12$)	31 and over	7.0	0.0	0.0
	6 to 30	17.0	1.4	0.3
	2 to 5	25.0	4.9	2.5
	1	51.0	93.7	97.1
	Total	100.0	100.0	99.9

* The data for Group I summarized in this table include only seven of the eight *S*s. The latency data for one *S* were excluded because they were so radically different from those of all other *S*s that they would have obscured the general trends of the other members of the group.

clearly apparent when Groups III and IV are contrasted than when Groups I and II are contrasted. To some extent this may be due to certain improvements in techniques with Groups III and IV, but the greater likelihood is that the differences in clarity of the results are due to certain other factors to be discussed later in this section. Among the improvements in technique involving Groups III and IV are the following:

1. In the first place, Groups III and IV were composed of split litters, animals of each group of which were matched for weight, age, sex, naïveté, and equivalence of environmental background. Groups I and II consisted of animals matched for age, sex, weight, naïveté, and synonymity of environmental background, but, of course, were not from the same litters, since Group II training did not begin until Group I training was at an end.

2. In the second place, Groups III and IV were given trials on the same days. This meant increased likelihood of equivalent handlings. Groups I and II, of course, were run several weeks apart on the corresponding trials.

3. In the third place, Groups III and IV received preliminary training. Groups I and II did not. Hence, it is difficult to know how much of the early part of the experiment for those animals (Groups I and II) was devoted to general exploration within the apparatus. One can be more certain that the animals in Groups III and IV who received preliminary

training were attentive to relevant stimulus cues in the early part of the training period.

4. In the fourth place, the sizes of the samples in Groups III and IV were larger than in I and II.

According to the Spence theory of discrimination learning, learning consists of the relatively simple process of developing differential strength among alternative stimulus-response connections, as determined by selective reinforcement and nonreinforcement. This concept emphasizes the frequency of stimulus-response-reinforcement sequences. Each occurrence of this sequence results, according to the theory, in building up cumulatively a certain amount of associative (habit) strength. Likewise, according to the theory, the stimulus-response-nonreinforcement sequence produces a cumulative building-up of inhibitory strength.

If this theory were valid, then each animal in Groups II and IV, respectively, since they received more reinforcements for responses to the positive stimuli and more nonreinforcements for responses to the negative stimuli than the poorest animals in Groups I and III, respectively (trained relationally to the same stimuli), should have developed, to a much greater extent than any animal in Group I or III, the appropriate habits involved in the solution to the discrimination problem. But such was not the case.

Despite the fact that Group II animals each received more positive (reinforced) trials and more negative (nonreinforced) trials to the appropriate stimuli than even the poorest animal in Group I, the behavior of some of the animals in Group II was not equal to the behavior of all of those in Group I. On being tested, five of the animals did attain the same criterion of "perfection" as the animals in Group I; two other animals, each of which made one error during the tests, practically reached the criterion. But one animal made errors on 10 per cent of the criterion runs, another made errors on 27.5 per cent of the criterion runs, and another made errors on 32.5 per cent of the criterion runs. This is clearly not equal to the "perfect" performance of every animal in Group I. However, because of the fact that, in the respects mentioned above, the procedure employed with Group I and II *S*s was less adequate than that employed with Groups III and IV, our attention should be concentrated upon a comparison of the latter groups. The evidence in regard to these groups is sharp and distinct. It clearly favors the gestalt interpretation. Although each animal in Group IV, the absolute stimulus group, received as many positive (reinforced) and as many negative (nonreinforced) trials as the poorest animal in Group III, the relational stimulus group, *none of the animals in Group IV even remotely approached the criterion of learning which all of the animals in Group III met.* One animal (Group IV, No. 8) obtained a "chance" score of 50 per cent. The animal accomplished this score by adopting a position habit from the very first test trial. During the training in the days prior to the test, the animal had responded with equal alacrity to either stimulus component by approaching it, regardless of the compartment in which the stimulus was exposed. (It did not matter whether the stimulus presented was negative or positive.) No other animal on the discrimination tests for Group IV achieved a score of over 45 per cent. The mean success of these animals was 32.5 per cent; and one animal (Group IV, No. 2), on being tested, selected the dimmer stimulus, which was the positive, only once in the 40 test trials, making an accuracy score of 2.5 per cent. Hence, it is apparent that the animals in Group IV, like the animals in Group II, displayed a strong preference for the brighter of the two lights. And this occurred despite the fact that the training was exactly the opposite in terms of the differential reinforcement and nonreinforcement of the same stimuli. Clearly the results are not those which the associationistic theorists would predict. The Spence theory would hold that, with such differential training, opposite specific habits for the two groups should have been developed, but obviously this was not the case.

B. *A Suggested Explanation for the Response Tendency to the Brighter Light in Both Absolute Groups*

In view of the fact that differences between Groups III and IV are obvious, the less obvious differences between Groups I and II demand explanation. Actually, of course, one cannot directly compare the first two groups with the last two groups because of at least one difference in procedure, viz., Groups III and IV received preliminary training whereas Groups I and II did not. However, the following is a suggested explanation:

For the absolute groups, as a consequence of partial reinforcement to light, a general light-going tendency, *but not a tendency to go to specific light,* was developed. The learning situation for these groups was one in which a single light was exposed, the other stimulus window being covered and apparently indistinguishable from the dark ground. The animals, therefore, learned to go to a spot of light. And this is what they continued to do when tested. Since two spots of light differing greatly in brightness were presented in the test, the rats reacted to the more prominent. Hence, the animals in both Groups II and IV, regardless of which stimulus they were trained to approach (the brighter or the dimmer), chose the brighter stimulus on being tested.

Hull, Spence, and other associationistic theorists would have predicted that animals with reinforcement for responses to a particular stimulus would develop the strength of the excitatory tendency (i.e., the positive tendency or tendency to evoke the approaching response) with regard to this stimulus; and with nonreinforcement for responses to another specific stimulus would develop the inhibitory tendency (i.e., the negative tendency) with regard to that stimulus. Hence, the animals in Group IV should have approached the dimmer stimulus and avoided the brighter stimulus as a consequence of their previous training. This was not the case. There was not even a trend to approach the positive stimulus. In fact, taken at face value, the evidence seems to indicate that the training had an adverse effect, for despite previous training toward the dimmer light, on exposure to both lights simultaneously, the animals in Group IV, like the animals in Group II, tended to approach the brighter.

C. *Discussion of Position Habits Developed and of Light Reactions During Tests of Absolute Groups*

In interpreting the results of tests involving the absolute groups, at least three facts must be considered.

1. The final testing situation involved the simultaneous presence of two stimuli, whereas prior to this time (i.e., during the training trials) the animals had been confronted by only a single stimulus at a time. It will be recalled that during the training period *S* approached each of the light stimuli at different times when individually presented. If during the final test *S* tended to react to both of the

Bureau Ednl. Psy. Research
DAVID HARE TRAINING COLLEGE
Dated ....
No. ....

simultaneously presented lights in the same manner as to these stimuli when individually presented, the situation would be one of confusion.

2. Tremendous differences existed in the intensity of the two lights; one light was physically several times as bright as the other. Hence, the animals might respond to only the most prominent (brightest) of the two, the other light being relatively "unnoticed."

3. As a consequence of differential reinforcement and nonreinforcement for the stimuli during the early trials of the final test situation (which permitted relational discrimination), modification of response tendencies on the *later* trials of the final test situation might have occurred.

In Group II tests, animals on initial trials tended to respond to the most prominent of the two lights and were always rewarded for such responses. Although the presence of two lights might have been confusing, the reinforcement for brightness responses tended to confirm this reaction tendency, i.e., the tendency to approach the brighter light. But even in this group two animals developed position habits. One *S* made all runs to the left on the first two test days, and on the third day made 8 of its 10 runs to the left; but on the fourth day it responded to the brighter light on each trial. The fourth day's responses might have been a consequence of the differential reinforcement and nonreinforcement of the first three days. Another *S* responded to the most prominent, i.e., the brighter light, on the first day's complement of test trials. However, on the second day, 9 of the 10 responses were again to the right; on the third day, 9 of 10 responses were again to the right, on the fourth day, 8 of 10 responses were to the right side. It is suggested that effective stimulation of the animal by two stimuli after the first day provided a situation of confusion responsible for the development of position habits.

In the case of Group IV, animals on initial trials responded to the most prominent of the two lights and were never rewarded for so doing. Thus, there was not reward, but a situation of disturbance, of constant frustration for those reactions. The presence of two lights, of course, might have produced an additional confusing factor. Under these conditions, behavior was modified, and position habits were established. Since the position habits were partially reinforced, whereas the previous responses to the brighter stimulus were not reinforced, they (the position habits) tended to persist. It is possible that these position habits, too, would have been resolved if sufficient practice were allowed in the relational situation; but this would have changed the situation from one of test to one of further training by a second method.

D. *Discussion of Transposition Test Results*

The results of transposition tests with Groups I and III, in which *S*s, after being trained to a standard pair of stimuli, were tested with the positive stimulus or the negative stimulus of an original (standard) pair and a new stimulus, still further positive or negative, indicate that the *relationships* between the stimuli and *not their absolute properties* are the basis for discrimination.

E. *Comment on the Difficulty of the Discrimination Problems*

In spite of the fact that a greater mean number of trials was required to learn the discrimination in which the brighter stimulus was positive as compared with the same pair of stimuli with the dimmer positive (the mean for Group I was 316.3; the mean for Group III was 309.2) it is suggested that the discrimination with the dimmer as positive is a more difficult one to learn. While the direct comparison of the first two groups and the last two groups is restricted by the differences in procedure employed, it is doubtful that such procedural differences, such as preliminary training, influenced the results, except in the rate of learning. It is probable that the preliminary training of Group III (see preliminary training, under "Procedure" in Section II) served to familiarize the animals with the learning apparatus so that general exploratory behavior was less necessary when the training trials, per se, started. Hence, since some familiarity with the apparatus was accomplished prior to the formal training, diffuse, exploratory behavior in the apparatus was reduced, and the total trials required for learning were correspondingly reduced.

F. *Suggestions for the Design of a Future Experiment to Obtain Further Data on this Problem*

Even though the results of this investigation, particularly the comparisons involving Groups III and IV are fairly clear, at least three suggestions may be made which the investigator believes would contribute to the evidence and yield even more obvious results in the same direction.

1. First of all, differences in physical intensity between the two stimuli should be reduced and differential thresholds of brightness for the rat should be determined in a preliminary investigation. Settings of the stimuli for training and testing in the experiment would then depend on the results of this preliminary investigation.

In the current investigation, stimuli that were obviously different from each other were employed. One stimulus was approximately seven times as bright as the other (in terms of physical units). Moody, in one investigation of the deermouse, found that the minimum differential threshold between brightness stimuli was 1 to 1.6 (14). It is likely that the ratio for the white rat is somewhere in the vicinity of this ratio.

There is, of course, little doubt, that if stimuli differ tremendously from each other so that they set up secondary responses and sensations, probably they could be learned in absolute terms. Hence, the problem should involve stimuli which differ from each other minimally rather than maximally.

That these two stimuli (used in the present investigation) were readily distinguishable from each other is obvious, for Group IV animals did respond in the test with a decided preference for the bright light. Thus, although they did not learn the solution to the discrimination problem, they did distinguish between the stimuli. If the differences between the stimuli were greatly reduced so that they were less readily distinguishable, probably there would be no differential reaction except with relational training.

2. Second, *Ss* should receive a considerable amount of handling prior to the beginning of the experiment to ensure tameness, and should have had considerable familiarity with the apparatus. Animals sophisticated in this way are easier to deal with than the more naive. It also seems reasonable to assume that such animals will learn most rapidly, since they are not so apt to attend to exploration and irrelevant stimuli at the beginning.

3. Third, the space between the stimulus windows should be reduced. In the present investigation, the internal edges of the two stimulus windows were 5½ inches apart. If these stimulus windows were moved nearer together so that the internal stimulus borders of each were only one or two inches apart from each other, this condition might provide a better opportunity for contrasting and comparing of the two stimuli when presented simultaneously; this condition, of course, according to advocates of associationistic theories, should have no particular effect upon the absolute stimuli individually presented.

G. *Interpretation of Essential Findings*

Absolute training in which no opportunity for direct and simultaneous comparison of the component stimuli exists is not equivalent in efficacy to relational training in which there is an opportunity for direct and simultaneous comparison of the component stimuli. When the term *comparison* is used here, we are not speaking of a higher mental process or a judgmental process, but merely implying that the organism can respond on the basis of an objective relationship.

The discriminatory response is not strictly a cumulative function of building up specific positive habits to specific stimuli as a consequence of a stimulus-response-reinforcement sequence, and a cumulative function of building up of inhibition as a consequence of a stimulus-response-nonreinforcement sequence.

The relationship between the members of a pair of stimuli simultaneously present has significance in "guiding" the behavior of the organism. Thus, the evidence offers support for gestalt-type concepts of discrimination learning and is opposed to associationistic theories such as those proposed by Hull and Spence.

Thus, if this research can be considered to embody processes fundamental to discrimination in general, learning of a discrimination problem is a function of the pattern of the stimuli; the learning process does not appear to be reducible to, nor can it be adequately discussed in terms of, separate and individual training to the absolute properties of simple stimuli. In terms of this particular research it seems that it is the relationship *between* the stimulus constituents of a pattern, rather than the individual properties *of* each stimulus constituent of the pattern, which is of significance in discrimination learning.

V. SUMMARY

The problem. The present investigation attempted to obtain evidence regarding the adequacy of training in discrimination problems under (*a*) relational conditions, in which both the negative and positive stimuli were simultaneously present, and (*b*) under absolute conditions in which only one stimulus, either the negative or the positive, was displayed (but not both simultaneously). Gestalt theorists propose that discrimination is a relational process. Many behavior theorists, such as Spence, hold that discrimination is essentially a function of the number of reinforcements and nonreinforcements following reactions to each stimulus, i.e., that discrimination is a matter of developing positive response tendencies (habit strength) and negative response tendencies (inhibitory strength) to each stimulus, and discrimination is reducible to differential training with each individual stimulus component separately in an absolute fashion.

Subjects employed. Four groups of female rats were employed.

Group I (N = 8) These animals were given relational training with the brighter stimulus window as positive.

Group II (N = 10) These animals were given absolute training with the brighter stimulus window as positive.

Group III. (N = 12) These animals were given relational training with the dimmer stimulus window as positive.

Group IV. (N = 12) These animals were given absolute training with the dimmer window as positive.

Procedure. The *S*s in Group I were trained until they reached the criterion of four successive days without error (ten trials per day). The *S*s in Group II were trained by the absolute method and given as many (or more) trials to each stimulus as were required by the poorest *S* in Group I to attain the criterion. Then Group II *S*s were tested with both stimuli present. The *S*s in Group III were trained until they reached the criterion of four successive days without error (ten trials per day). The *S*s in

Group IV were trained by the absolute method and given as many (or more) trials to each stimulus as were required by the poorest *S* in Group III to attain the criterion. Then Group IV *S*s were tested with both stimuli present.

*Results. All S*s in the relational groups, Groups I and III, reached the learning criterion. The *S*s in Group II, on being tested by simultaneous presentation of the two stimuli, exhibited a strong preference to approach the brighter of the two lights. The *S*s in Group IV, on being similarly tested, also exhibited a strong tendency to approach the brighter of the two lights, despite the fact that they (Group IV) had been trained to go to the dimmer stimulus; none of them even remotely approached the learning criterion. Hence, regardless of whether animals subjected to the absolute method were trained to go to the brighter or to the dimmer stimulus, when both stimuli were presented simultaneously, all, with the exception of one animal who displayed a position response, preferred the brighter.

The results are interpreted to support the gestalt concept, i.e., that discrimination is based upon the "comparative" (or relational) aspects of the immediate component stimuli of the configuration, and that training upon the component stimuli separately is inadequate, at least less adequate, in producing efficacious discriminatory reactions.

BIBLIOGRAPHY

1. BINGHAM, H. C. Size and form perception in Gallus Domesticus. *J. anim. Behav.*, 1913, 3, 65-113.
2. CASTEEL, D. B. The discriminative ability of the painted turtle. *J. anim. Behav.*, 1911, 1, 1-28.
3. COBURN, C. A. The behavior of the crow, *Corvus Americanus Aud.*, *J. anim. Behav.*, 1914, 4, 185-201.
4. GULLIKSEN, H., & WOLFLE, D. L. A theory of learning and transfer: I. *Psychometrika*, 1938, 3, 127-149.
5. GULLIKSEN, H., & WOLFLE, D. L. A theory of learning and transfer: II. *Psychometrika*, 1938, 3, 225-251.
6. GUNDLACH, R. H., & HERRINGTON, G. B., JR. The problem of relative and absolute transfer of discrimination. *J. comp. Psychol.*, 1933, 16, 199-206.
7. HULL, C. L. *Principles of behavior*. New York: Appleton-Century-Crofts, 1943.
8. KINNAMAN, A. J. Mental life of two macacus rhesus monkeys in captivity. *Amer. J. Psychol.*, 1902, 13, 98-148, 173-218.
9. KOHLER, W. Aus der Anthropoidenstation auf Teneriffa. IV. Nachweis einfacher Strukturfunktionen beim Schimpansen und beim Haushuhn: Über eine neue Methode zur Untersuchung des bunten Farbensystems, *Abh. preuss. Akad. Wiss.*, Berlin, 1918, 1-101.
10. KRECHEVSKY, I. Hypotheses in rats. *Psychol. Rev.*, 1932, 39, 516-532.
11. LACHMAN, S. J. A theoretical and experimental investigation of absolute and relational stimulus training in discrimination learning. Unpublished doctor's dissertation, Univer. of Michigan, 1951.
12. LASHLEY, K. S. *Brain mechanisms and intelligence*. Chicago: Univer. of Chicago Press, 1929.
13. LASHLEY, K. S., & WADE, MARJORIE. The Pavlovian theory of generalization. *Psychol. Rev.*, 1946, 53, 72-87.
14. MOODY, P. A. Brightness vision in the deer-mouse, *Peromyscus Maniculatus Gracilis*. *J. exp. Zool.*, 1929, 52, 367-405.
15. MUNN, N. L. *Handbook of psychological research on the rat*. New York: Houghton Mifflin, 1950.
16. SPENCE, K. W. The nature of discrimination learning in animals. *Psychol. Rev.*, 1936, 43, 427-449.
17. SPENCE, K. W. The differential response in animals to stimuli varying within a single dimension. *Psychol. Rev.*, 1937, 44, 430-444.
18. TAYLOR, H. A. A study of configurational learning. *J. comp. Psychol.* 1932, 13, 19-26.
19. WARDEN, C. J., & ROWLEY, J. B. The discrimination of absolute versus brightness in the ring dove, *Turtur risorius*. *J. comp. Psychol.*, 1929, 9, 317-337.
20. WARDEN, C. J., & WINSLOW, C. N. The discrimination of absolute versus relative size in the ring dove, *Turtur risorius*. *J. genet. Psychol.*, 1931, 39, 328-341.

(Accepted for publication October 7, 1955)

Vol. 70, No. 7 Whole No. 414, 1956

Psychological Monographs: General and Applied

The Relation Between Possession of Task-Relevant Information and Attempts to Lead[1]

John K. Hemphill, Pauline N. Pepinsky, Reuben N. Shevitz,
William E. Jaynes, and Charlotte A. Christner[2, 3]

Personnel Research Board, The Ohio State University

I. INTRODUCTION

A. General Statement of the Problem

The present experiment is the first of a projected series concerned with the prediction of individual attempts at leadership in small groups. Specifically, the study to be reported is an inquiry into the extent to which attempted leadership acts are a function of the possession of task-relevant information. "Attempted leadership acts" are acts accompanied by an intent to initiate structure-in-interaction during the process of solving a mutual problem. The central hypothesis under test is drawn from a theory of leadership in small groups proposed by Hemphill (23, Appendix A). In general, the current project is restricted to the problem of defining and describing the motivational factors underlying individual attempts to lead. Whether such attempts are "successful" (i.e., accepted or followed by the group members) or "effective" (i.e., do, in fact, lead to correct problem solution) are at this stage important but irrelevant questions.

B. Relevant Concepts and Postulates from a Proposed Theory of Leadership in Small Groups

The hypothesis tested in the present experiment makes use of selected con-

---

[1] This study was conducted under Contract N6ori-17 T.O. III NR 171 123 between the Office of Naval Research and the Research Foundation of the Ohio State University. The opinions presented are those of the authors and should not be regarded as having the endorsement of the Department of the Navy. The research is a part of The Ohio State Leadership Studies, which were initiated in 1946 by Carroll L. Shartle and are planned as a ten-year program of research on leadership and administrative problems in military, business, industrial, educational, and civilian governmental organizations.

This paper is based on a technical report submitted to the Office of Naval Research which bears a similar title (see 23). Certain materials, e.g., a complete statement of the theory and sample forms, have not been included in the present paper in order to reduce printing costs. The original report is available through interlibrary loan from The Ohio State University.

[2] The authors are grateful to Dr. Harold Garfinkel, now of the Department of Sociology, University of California at Los Angeles, for his invaluable advice in the initial planning stages of the experiment. They wish to thank Dr. Delos D. Wickens of the Department of Psychology, The Ohio State University, for his cooperation in securing subjects.

Acknowledgment is due also to Mr. Howard J. Ehrlich for his service as an observer, to the students who served as subjects in the experiment, and to Mrs. Arlene Bargar, Mrs. Ruth Ann Young, and Miss Evelyn Harris for their conscientious clerical assistance.

[3] John K. Hemphill is now with the Educational Testing Service, Princeton, New Jersey; Reuben N. Shevitz is with E. I. Lilly & Company; William E. Jaynes is with the United States Army.

No. 6 wanting

Bureau Ednl. Psy. Research
DAVID HARE TRAINING COLLEGE
Dated
Acc. No

cepts central to the theory and is derived from theorems that stipulate the variables upon which attempts to lead are contingent. The key concepts employed are those of "mutual problem," "interaction," "structure-in-interaction," and "the attempted leadership act." These concepts are defined as follows:

1. "A *mutual problem* is a situation involving two or more individuals, each of whom perceives the situation as dissatisfying, and that requires some behavior on the part of the other if either (any) is to be satisfied" (23, p. A-2).

2. "Interaction is defined as a series of reciprocal acts involving two persons, where the intentions governing the acts are influenced by acts, or expectations of the acts, of the other" (23, pp. A-6, A-8).

3. "Structure-in-interaction is defined as a consistency in behavior occurring during interaction that permits the prediction of the behavior that will occur in future interaction" (23, p. A-8).

4. "*Attempted* leadership acts are acts accompanied by an *intention* of initiating structure-in-interaction of others during the process of mutual problem solution" (23, p. A-18).

The theory holds that "the motivational variables that determine whether a given individual will attempt a given leadership act lie in two major areas. The first area involves need-deprivation variables. . . . The second area involves cognitive estimates or expectations of the individual concerning the probable consequence of the act of leadership he is considering attempting" (23, p. A-26). Within the second area "the possession of relevant information (perhaps not available to other group members)" is specified as a variable that contributes to the individual's ability "to follow clearly the probable consequence of the act under consideration" (23, p. A-27). Under this circumstance he may attach a relatively high probability of effectiveness to an act that he is considering. "On the other hand, an individual will tend to estimate a low probability of effectiveness of a course of action that appears novel with respect to his previous experience and about which he is unclear and confused as to its likely consequences" (23, p. A-27).

C. RELEVANT RESEARCH AND THEORY

Previous research on leadership and group behavior has been summarized and discussed elsewhere (14, 34, 26, 18), and Roseborough (31) has prepared a recent survey of experimental studies with small groups. It will not be necessary, therefore, to review here the large body of accumulated research, but it is appropriate to point to certain characteristics of other studies that have particular pertinence in the present context.

First, and of special significance, is the virtually complete lack of attention given to the problem of attempted leadership. A number of studies involving an apparently similar notion have had to do with the concept of "emergent" leadership as occurring in a "leaderless" group situation—e.g., see Ansbacher (1), Gibb (16), Bass (3), and Carter, Haythorn, Shriver, and Lanzetta (9). It is clear, however, from the published descriptions of the operations by which the behavior of emergent leaders has been assessed, as opposed to those used to measure attempts at leadership, that the terms *emergent* and *attempted* leadership are not identical. The difference lies in the fact that, in the former instance, leadership behavior, however defined, has been measured in retrospect, in the sense that only after a "leader" has emerged, does his behavior become a phenomenon of interest to the observer. In the latter case, the focus is upon conditions and variables that may motivate an individual simply to *try* to lead, without any regard for whether he does later, or ever, "emerge" as a leader, either in the eyes of the group members or of any outsiders.

Nevertheless, it should be granted that an undetermined proportion of an individual's attempts to lead may be at least partly responsible for his subsequent recognition as a leader. It is plausible, therefore, to assume that there is some degree of communality between what would be viewed by one observer as leadership behavior and by another as attempted leadership only. But in each case a different question is raised about that behavior: on the one hand the problem is to determine the factors related to different *leader* behaviors (taking for granted that a leader has been identified); on the other hand, the problem is to find out what variables are associated with *any* group member's attempts to lead.

An exception is the recently published monograph by Bass, Wurster, Doll, and Clair (4).

This study of situational and personality factors in leadership among sorority women makes use of a concept of attempted leadership. In this instance, however, attempted leadership is defined simply as the amount of participation and is treated as one of a series of measures obtained through employing the leaderless group discussion as a situational test. In view of the more restrictive definition of attempted leadership formulated by Hemphill, it is difficult to make a meaningful comparison between the present research and the investigation by Bass and his associates.

A second point of relevance to the formulation of the hypothesis tested in the current investigation is the recurrent finding, commented upon by Jenkins that "leaders in a particular field need and tend to possess superior general or technical competence or knowledge in that area" (26, p. 75). As Gibb (15, p. 22) has noted, Caldwell reported as early as 1920 that his study of nominated high school leaders had led him to conclude that one of the most important characteristics ascribed to these individuals was the possession of "practical knowledge relative to the situations for which they were chosen as leaders" (6). Although attention has been given to the possession of special knowledge or training as an attribute of *acknowledged* leaders, the question does not appear to have been asked as to whether, given an individual who possesses special training or information, it then can be predicted that he will *attempt* to lead.

A third frequently stated conclusion, congruent with the one above and also cited by Jenkins, is that "leadership is specific to the particular situation under investigation" (26, p. 75), also see Hemphill (21). Recently, some qualifications of this categorical statement have been introduced by the studies of Carter, Haythorn, Meirowitz, and Lanzetta (8) and Bell and French (5). But their findings serve not so much to contradict the specificity hypothesis as to underscore the complexity of the phenomona under scrutiny. The fact remains that the degree of specificity that has seemed to characterize the behavior of leaders and the variables associated with leadership has been one of the most difficult obstacles to surmount in the attempt to derive systematic generalizations in the field of leadership study.

Recognition of this state of affairs has led to an expressed need for more knowledge about the effects of particular situational variables upon group behavior. Roseborough says, "Such knowledge is vital to every small-group study since the results are conditional on these factors. Too often situational variables have been uncontrolled and unspecified in the design of experiments" (31, p. 297). And Heise and Miller have urged the importance of defining the significant dimensions of the problem itself, citing as possibilities "the degree to which the problem is structured . . . the uniqueness of the solution, *the initial distribution of information* (italics ours), the number of decisions that some one must make," etc. (20, pp. 334-335).

With respect to the conceptual views underlying current research on leadership in small groups, both Katz (28) and Smith (33) have remarked upon the shift from the earlier concern with the "traits of leaders" to a relational concept of leadership, or what Gibb has referred to as an "interactional theory" (15). Brief mention will be given to two such efforts toward systematic formulation of some principles of group behavior that seem to have particular aspects in common with the theory proposed by Hemphill. The first is Deutsch's theory of cooperation and competition (10, 11, 12). Specifically, he has linked the concept of group with the concept of cooperation in that the cooperation situation is viewed as equivalent to the sociological group. "A sociological group exists (has unity) to the extent that the individuals composing it are pursuing promotively interdependent goals" (11, p. 150). The phrase "promotively interdependent goals" implies that the attainment of any group member's goal is dependent upon every other group member attaining his goal. There appears to be a close analogy between Hemphill's concept of the activity of group members engaged in solving a "mutual" problem and Deutsch's concept of the sociological group defined by the pursuit of promotively interdependent goals. Increased knowledge of the conditions and variables associated with cooperative behavior in groups may have implications for the study of attempted leadership in groups working toward the solution of mutual problems.

A recent paper of Newcomb's (30) sketches a theory of communicative behavior, which reveals some similarity of conceptualization to Hemphill's formulation. Newcomb has defined the communicative act, the unit of interaction, as "any observable behavior by which information, consisting of discriminative stimuli, is transmitted from a human source to a human recipient" (30, p. 141). It is his proposal that human social behavior "be studied in terms of the conditions and consequences of varying communicative acts" (30, p. 141).

It is possible that attempted leadership acts, if regarded as a class of communicative acts, could be accounted for in such a system as Newcomb suggests. Thus, in the case of the approaches to the study of group behavior proposed by Deutsch, and by Newcomb, certain parallels to the investigation of attempted leadership are discernible.

D. THE HYPOTHESIS TO BE TESTED

The central hypothesis tested in this experiment is drawn from the selected portions of the theory given in Section I, B. The postulated relationships have led to the formulation of the proposition that the possession of relevant information, which is not available to others, will be significantly related to the frequency with which a given individual will attempt acts of leadership during the course of the problem-solving activity of a given set of persons. Stated in null form, the specific hypothesis submitted to experimental test is as follows: in a mutual problem-solving situation the frequency of attempts to lead by those individuals who possess task-relevant information is not significantly different from the frequency of leadership acts attempted by individuals who possess nonrelevant information.

II. PROCEDURE

A. GENERAL DESCRIPTION OF THE EXPERIMENTAL DESIGN

The experiment was conducted according to a factorial design. The major dependent variable was the number of attempted leadership acts that were tallied for each of four group-members during the time that they worked on each of four tasks. A time limit of 15 minutes was set for each task and the groups proceeded from one task to the next without any appreciable delay. The principal independent variable was the task-relevant information supplied to each member during a training period that preceded the experimental sessions.

During the training session each group member was provided with a booklet of information, which he was permitted to study for forty minutes. Each of these booklets, made up and coded in advance, contained four sets of informational material of approximately equal length, but of varied relevance to the particular tasks. For each task two sets of information were prepared, one clearly more relevant and the other less relevant to the solution of the problem. The prearranged composition of the booklets was such that one group member's booklet might contain three sets of irrelevant and one set of relevant information, while another's contained two sets of relevant and two of irrelevant information, etc.

The four tasks used in the experiment were an Assembly task (Task A), a Strategy problem (Task S), a Construction task (Task C), and a Discussion problem (Task D). Sixteen groups of four members each served as subjects in the experiment. They began their first task as unacquainted individuals and as members of "leaderless" groups. Table 1 illustrates the basic design. The order in which the tasks were presented and the number of individuals possessing relevant information are given for the first eight sessions, or one-half the total number. During the remaining eight sessions an experimental replication was provided by repeating the basic design.

The subjects were selected from among the male university students who were enrolled in an elementary psychology course. The selection procedure involved choosing volunteers who were found, on the basis of a screening test, to be relatively naive in respect to the types of task used in the experiment.

At the beginning of the experimental sessions the subjects were told that they would be in competition with other groups for monetary prizes to be awarded

TABLE 1

BASIC DESIGN OF THE EXPERIMENT

Group I						Group II						Group III						Group IV					
Trial	Task	Individual				Trial	Task	Individual				Trial	Task	Individual				Trial	Task	Individual			
		1	2	3	4			5	6	7	8			9	10	11	12			13	14	15	16
1st	A	N	N	N	N	1st	C	R	R	R	R	1st	S	N	N	N	N	1st	D	R	R	R	R
2nd	S	R	R	R	R	2nd	D	N	N	N	N	2nd	A	R	R	R	R	2nd	C	N	N	N	N
3rd	D	R	N	N	N	3rd	A	R	N	N	N	3rd	C	R	N	N	N	3rd	S	R	N	N	N
4th	C	R	R	R	N	4th	S	R	R	R	N	4th	D	R	R	R	N	4th	A	R	R	R	N

Group V						Group VI						Group VII						Group VIII					
Trial	Task	Individual				Trial	Task	Individual				Trial	Task	Individual				Trial	Task	Individual			
		17	18	19	20			21	22	23	24			25	26	27	28			29	30	31	32
1st	S	R	R	R	R	1st	D	N	N	N	N	1st	A	R	R	R	R	1st	C	N	N	N	N
2nd	A	N	N	N	N	2nd	C	R	R	R	R	2nd	S	N	N	N	N	2nd	D	R	R	R	R
3rd	C	R	N	N	N	3rd	S	R	N	N	N	3rd	D	R	N	N	N	3rd	A	R	N	N	N
4th	D	R	R	R	N	4th	A	R	R	R	N	4th	C	R	R	R	N	4th	S	R	R	R	N

Tasks:
A. Assembly problem (Electronic circuit)
S. Strategy problem (Peggity game)
C. Construction problem (The Toy, airplane design)
D. Discussion problem (Election results)

Information:
N = possess information *not relevant*
R = Possess information *relevant*

on the basis of their task performance. The four groups with the best task performance were to be given cash awards.

No restriction was placed on the subjects' freedom to move about and converse with each other. The fact that subjects could communicate freely with one another precluded the possible differential effect upon the incidence of attempts to lead of position within a communicative network. See Leavitt (29, p. 46).

B. THE DEPENDENT VARIABLE

1. Observational Procedure

As noted above, the observed frequency of attempted leadership acts constituted the dependent variable in this experiment. The particular procedures followed by the observers were governed by the requirement that their observations and inferences be linked directly to the theoretical definition of the attempted leadership act as "an act accompanied by an intent to initiate structure-in-interaction of others during the process of mutual problem solution."

Reference to such systems of categorization of observed group behavior as those developed by Bales (2) and Carter, Haythorn, Meirowitz, and Lanzetta (7) suggested that certain of their categories might include behavior similarly codified. Specifically, Bales's fourth category ("gives suggestion, direction, implying autonomy for others" (2, pp. 181 ff.), or Carter's categories 25 and 26 ("proposes course of action for self" and "proposes course of action for others" (7, p. 259) might plausibly be expected to subsume many "attempted leadership acts."

But without a subsidiary methodological study, which is outside the scope of this experiment, it would not be possible to say to what extent these labels do have common behavioral referents. As Heyns and Zander have emphasized, a decision about the applicability of a category system "depends upon the purposes of the investigator and the theoretical framework within which he is working" (25).

Accordingly, in an attempt to achieve maximum clarity and precision, an Observer's Manual was developed especially for use in the current research. From the theoretical definition of the attempted leadership act a set of procedural rules was derived that specified those acts clearly included in the system and to be scored, and those acts clearly excluded and not to be scored. Thus, the category system comprised what Guetzkow has called a "sieve code" (17, p. 48) in that only the behaviors falling into a single major category were tabulated; all other behavior was ignored. The manual also outlined procedures for categorizing acts that were "marginal" in the sense that scoring them required relatively complicated inferences on the part of the observer. Three trained observers independently tabulated attempted leadership acts.

One of these observers had not participated in the earlier development of the scoring rules, was unfamiliar with the underlying theory, and was trained solely by reference to the manual in its final form. This observer was introduced in order to determine whether the manual had made the concept of the attempted leadership act sufficiently communicable to yield an adequately reliable measure.

2. *Reliability of Observations*

There is no generally satisfactory solution to the problem of assessing agreement among observers in small-group experimentation. For the purpose of this experiment, the intraclass correlation (r') technique (27, p. 230 ff.), appeared to provide an adequate means of measuring interobserver consistency. This technique was applied to obtain, first, a measure of over-all agreement within the set of three trained observers, and, second, separate measures of agreement within each of the three pairs of observers. The scores used in this analysis were the distributions of the total number of attempted leadership acts assigned by each observer to each of the four individual subjects during every fifteen-minute period of the sixteen sessions. Thus, there were 256 separate scores for each observer. The obtained correlations are presented in Table 2.

Particularly in view of the fact that the observers were required to draw inferences with respect to the presence

TABLE 2
INTRACLASS CORRELATIONS AMONG TRAINED OBSERVERS' TALLIES OF ATTEMPTED LEADERSHIP ACTS

Observer set	Source of variation	Sum of squares	*df*	*F*	r'
A+B+C	Between sets Within sets	25,388.69 3,250.67	255 512	15.69**	.83
A+B	Between pairs Within pairs	19,171.38 1,254.00	255 256	15.34**	.88
A+C	Between pairs Within pairs	17,428.37 1,870.00	255 256	9.36**	.81
B+C	Between pairs Within pairs	15,391.88 1,752.50	255 256	8.81**	.80

** Significant at .01 level.

of an "intent" on the part of the group members, an over-all correlation of the magnitude of .83 was considered adequate for the purpose at hand. The highest intrapair correlation, .88, as might have been expected, occurred between the two observers who had participated in the design of the experiment and the development of the manual, and who were most familiar with the theoretical framework. Nevertheless, there was only a relatively small and obviously not significant difference between this correlation and the separate measures computed for each of these observers paired with the third observer, "C," trained with reference to the manual only. These correlations of .81 and .80 supported the contention that the concept of the attempted leadership act as defined theoretically could be given behavioral referents subject to reliable observation.

3. The Attempted Leadership Act and a Naive Concept of Leadership

For eleven of the sixteen sessions a fifth subject was available. The fifth subject was employed as a naive observer and was instructed simply to rate the group members on "amount of leadership." He worked behind a one-way vision mirror and made a judgment every 5 minutes.

Product-moment correlations were computed between (*a*) the sum of the naive observers' ratings for each of the four subjects during every 15-minute period of the eleven sessions, and (*b*) the total of the attempted leadership acts ascribed to each subject by the trained observers during the same time periods. The correlation obtained was .47 ($N = 176$). It would appear, therefore, that, while the experimenter's concept of attempted leadership is not entirely divorced from the concept of leadership held by the layman, the two notions show only a slight correspondence. An important qualification to be noted is that the rigorous restriction of the experimenter's concern to attempted leadership acts, without regard for their "success" or "effectiveness," had an unknown and possibly significant effect upon this measure of association.

C. THE EXPERIMENTAL CONTROLS

Certain methods of experimental control were introduced to insure the experimental manipulation of relevant information as the independent variable and to restrict the influence of extraneous variables. These controls were introduced in several ways: (*a*) in the development and selection of the experimental tasks to conform with certain specified criteria; (*b*) in the selection of subjects who were without previous knowledge directly relevant to the tasks; (*c*) in the employment of a pre-experimental training session held under controlled conditions to provide specified individuals with information relevant to particular tasks; (*d*) in the establishment of an external judgment criterion used to test the discriminability of the "relevant" and "irrelevant" sets of information; and (*e*) in the use of pre- and postexperimental testing procedures that provided a means both of relating amount of information to task performance and of determining the extent to which information was unevenly distributed throughout the experiment.

1. The Tasks

Four experimental tasks were devised, the characteristics of which were designed to facilitate the manipulation of task-relevant information and to maximize the likelihood that attempts to lead would be elicited:

a. Assembly task. The subjects were required to assemble a simple electronic circuit to conform to a prescribed schematic diagram. The materials included a box containing a large number of assorted components (resistors, condensers, coils, etc.) and an 18″ × 24″ sheet of cardboard upon which were drawn the schematic diagram and a set of squares. These squares corresponded to the elements of the schematic diagram, and were connected by straight lines, which represented the wires connecting the component parts. The subjects were told to assemble the circuit by selecting the correct parts from the box and placing them in the proper squares.

b. Strategy problem. This task involved participating in a game of peggity, which is played by inserting pegs into rows of holes on a large pegboard, with the objective of scoring by getting eight pegs in an unbroken row. The four group members played as a team against the experimenter, who played according to a fixed set of rules. Four pegs were placed on each move, each group member placing one. The team was permitted to discuss its moves before making them, but was required to reach agreement about the placement of all four pegs before each play. The team initiated the play by having the first move, and play continued until the expiration of the time limit.

c. Construction task. The group was assigned the task of constructing a model airplane from a set of commercially available materials called The Toy. Included in the materials were large square and triangular panels made of heavy colored paper, a number of wooden dowels, and a supply of pipe cleaners used to fasten the parts together. The group was instructed to build a model large enough for a four-year-old child to get inside, and was told that its design would be scored against one given in the manufacturer's blueprint.

d. Discussion problem. This problem required the group to arrive at an estimate of the total number of votes cast for the Republican presidential candidate in the state of Oregon in 1940. The group was asked to deliberate and arrive at a consensus approximating the official figures as closely as possible. A list of four facts about Oregon was made available to the groups.

2. Checks on Relevant and Irrelevant Information

In order to determine whether the prepared sets of information could be distinguished, on other than an a priori basis, as "relevant" and "irrelevant" to the tasks, another kind of check was introduced. After being provided with brief descriptions of the tasks, six judges made an independent assignment of each of these sets of written material to either the "relevant" or "irrelevant" category. By application of the chi-square technique, using the exact treatment appropriate for a two-by-two table as given by Fisher (13, pp. 100-102), a probability of .001 was obtained for three of the tasks, and of .04 for the fourth. Thus, for three tasks, in only one instance in a thousand would such "accuracy" in sorting be expected to occur by chance alone. Failure to categorize "accurately" by one judge, as occurred with Task A (Assembly), would be predicted by chance in four instances in a hundred. This appraisal supported the assumption that the information would be distributed in accordance with the experimental design.

3. Task Performance

The relation of performance to possession of relevant or irrelevant information was obtained by testing the significance of differences between the performance scores of those group members who possessed relevant information and those who did not. For this purpose, since the nature of the tasks had prevented the determination of individual performance scores, the group performance scores were assigned to individual members.

The group's score was treated as the score received by each subject, and t values were computed to test the significance of the difference between the means of the scores of the individuals who had relevant information and those who had irrelevant information. The results of these tests for each of the four tasks are shown in Table 3. On two tasks, the Assembly task and the Construction task, the individuals who possessed relevant information were members of groups whose performance scores were significantly higher than those of

TABLE 3

MEAN TASK PERFORMANCE OF SUBJECTS POSSESSING RELEVANT AND NONRELEVANT INFORMATION ($N_1 = N_2 = 32$)

Task	Relevant		Non-relevant		t
	M	σ	M	σ	
Assembly	6.22	2.14	2.16	2.52	6.88**
Strategy	2.94	1.88	3.09	2.25	− .29
Construction	6.59	1.81	4.53	2.20	4.12**
Discussion	3.53	2.60	4.84	2.98	−1.82

** Significant at .01 level.

individuals possessing irrelevant information. Although significant differences were not obtained with respect to the Strategy and Discussion tasks, the difference found in the case of the Discussion task did approach significance and occurred in a reversed direction.

While this analysis had the advantage of utilizing all the available data, it had an obvious disadvantage in that, on two of the problems, each group contained some members who had relevant information and some who did not. Thus some group scores were assigned both to individuals who possessed relevant information and to individuals who possessed irrelevant information. Furthermore, this procedure combined scores for individuals who were members of groups that were run under different information conditions; the effects on performance of possession of relevant information might have varied with the number of group members who possessed such information. Accordingly, the same sort of analysis was made again, using only the first two trials in which the members all had received the same (i.e., either all relevant or all irrelevant) information. Except in the case of the Discussion problem, the results of this additional analysis did not differ substantially from those reported in Table 3. The t value (3.87) obtained for this task is significant at the .01 level. The direction of the difference indicates that groups who had irrelevant information scored higher (i.e., performed better) than groups who had relevant information.

4. Distribution of Information

Since the major hypothesis was concerned with relevant information as it was possessed by different individuals engaged in solving different tasks, it was important to determine whether such information actually was unequally distributed at the outset and remained so at the end of the experimental session. Accordingly, two tests were administered to all group members: one, a reading comprehension test, was given immediately following the training session; and the other, a retention test, was given immediately following the experimental session. These two instruments, composed of multiple-choice items, were designed to measure the relative amount of information acquired and retained by the individual subjects.

An analysis of the differences in scores on the comprehension (pre-experimental) and retention (postexperimental) tests showed that, for all tasks, those who had received relevant information scored significantly higher than those who received irrelevant information. Table 4 shows the means and standard deviations of scores received by individuals in the two categories and the t's obtained when the differences between these categories were tested.

We may interpret the results of these information tests as indicating that the individuals who had been given the relevant information to read possessed superior knowledge both at the beginning and at the end of the experimental session.

TABLE 4

MEANS OF THE TWO GROUPS FOR RELEVANT ITEMS ON COMPREHENSION TEST, RELEVANT ITEMS ON RETENTION TEST, AND NONRELEVANT ITEMS ON COMPREHENSION TEST

($N_1 = N_2 = 32$)

Test	Task	Relevant		Nonrelevant		
		M	σ	M	σ	t
Comprehension Relevant	Assembly	2.78	1.07	1.72	1.17	3.79**
	Strategy	2.94	1.44	1.13	1.04	5.84**
	Construction	3.09	1.28	1.28	1.14	6.03**
	Discussion	2.84	.98	1.31	1.00	6.12**
Retention Relevant	Assembly	3.28	1.11	2.31	.93	3.73**
	Strategy	3.34	1.43	2.13	1.16	3.67**
	Construction	2.97	1.18	1.91	1.06	3.79**
	Discussion	2.91	1.25	1.34	.66	6.28**
Comprehension Nonrelevant	Assembly	1.84	1.17	2.97	1.45	−3.42**
	Strategy	.91	.82	2.69	1.23	−6.59**
	Construction	1.88	.94	4.22	.94	−9.75**
	Discussion	1.28	.89	2.72	1.22	−5.33**

** Significant at .01 level.

It appears also, that there was improvement in the test scores from the initial test (reading comprehension) to the final test (retention). In the cases of the Assembly task and the Strategy task *t* tests of the significance of the increases in test scores yielded values of 3.55 and 3.53 respectively (.01 level). In the cases of the Construction task and the Discussion task the *t*-test values were 1.22 and .27, which are not statistically significant. If all four tasks are taken together in one test of the significance of the increase in scores, a *t* value of 4.13 is obtained, indicating a significant improvement in knowledge about the test items after the experimental session. The higher scores on the retention test cannot be attributed to the difficulty level of the items in the tests, since the items of relevant information used in the reading comprehension test and the retention test for the subjects tested during the first half of the experiment were interchanged between the two tests during the second half of the experiment.

In summary, controls instituted through the factorial design and the experimental procedures operated in the following major ways:

1. Controls imposed by the observers' procedure assured the effective measurement of the dependent variable, attempted leadership acts.

2. Selection of novel tasks and procedures designed to eliminate subjects who possessed prior task-relevant information assured that information would be distributed in accord with the factorial design.

3. The classification by six independent judges of the information materials into a relevant or irrelevant category supported the assumption that information would be distributed initially in conformance with the factorial design.

4. Comprehension and retention tests administered before and after the experimental session indicated that more relevant information was possessed by subjects who had received relevant articles than by those who had received irrelevant articles.

5. An analysis of performance scores

indicated that for two tasks, Assembly and Construction, subjects possessing relevant information were members of groups making significantly higher performance scores than groups containing members with nonrelevant information.

6. The factorial design permitted the removal or control of variables such as task order, type of task, and individual differences in "leadership ability" in determining the effects of possession of information on attempts to lead.

III. RESULTS

A. The Data

In the course of the study three major classes of data were collected: (*a*) group performance scores for each of the four tasks and for the sixteen groups, (*b*) individual scores on information tests administered before and after the experimental sessions, and (*c*) frequency distributions of tallies of individual attempts to lead under the various experimental conditions. Analyses of the first two sets of data, to determine in one instance the relation between task performance and possession of relevant information, and in the second instance to check upon the manipulation of the independent variable, have already been reported (see Section II, C, 3, 4). The third class, however, comprising observers' tallies of attempted leadership acts, constituted the data of primary interest. The average number of attempts to lead during a 15-minute work period, as recorded by the three observers for each individual, ranged from 0.0 to 36.7. It is with reference to this particular set of data that we have examined the principal hypothesis of the study; namely, that an individual will attempt acts of leadership with a greater frequency when he does, than when he does not, possess task-relevant information. Through application of the analysis of variance, the effects of the experimental variables and the effects of their interaction on attempted leadership scores have been studied in detail. The obtained results will be presented in this section.

Before the analysis could be undertaken, it was necessary to determine whether two assumptions underlying the use of the analysis of variance could be met. In respect to the first assumption of normality of distribution, inspection of the raw data revealed a marked skewness with disproportionate representation in the low frequencies. By using as a transformation the square root of the sum of the attempted leadership acts noted by the three observers, a distribution not significantly different from the normal curve was obtained ($p > .30$). The integer values of the transformed scores then were used as a measure of an individual's attempted leadership acts. These attempted leadership "scores" comprised the basic data for all the subsequent analyses. In regard to the second underlying assumption of the analysis of variance, i.e., homogeneity of subclass variance, Bartlett's test (27, pp. 83-85) yielded a probability of .89, supporting the assumption of homogeneity of variance within the 128 pairs of scores.

The variance analysis focused upon the assessment of group differences, task differences, and possession of information, all treated as main sources of variation. It also was possible to examine the first-order interactions of these potential sources of variation. In addition, two subpartitions of the variance associated with group differences were possible. One of the subpartitions separated those groups that began the sequence of four tasks with relevant information from those that began the sequence without relevant information. The second partition was based upon the four different orders in which the tasks were presented to the groups.

B. Analysis

1. The Analysis of Variance

Table 5 presents the analysis of the variance of the attempted leadership scores in which tasks, groups (initial information and task sequences), and information represent the partitions by main potential sources of variation. In the original tests of significance (F_1), the interactions of Group × Tasks and Information × Tasks were found to be not significant. Accordingly, the sums of squares associated with these sources were combined with the residual to provide a new error term for a second series of tests of significance (F_2). The analysis is discussed in terms of the findings obtained from the second series of tests of significance.

From Table 5 we see that the variance sources, task differences and group differences, have a significant effect upon the attempted leadership scores. The effect of information on the attempted leadership scores appears indirectly in first- and second-order interactions involving group differences.

a. Task differences. The mean attempted leadership scores for the four tasks were as follows: Assembly, 2.67; Strategy, 5.19; Construction, 3.73; and Discussion, 2.77. The Strategy task elicited by far the largest number of attempted leadership acts. The fact that this task required group consensus on each of a series of moves fostered vigor-

TABLE 5

Analysis of Variation in Attempted Leadership Scores Associated with Group Differences (Initial Information and Task Sequences), Task Differences, and Relevance of Information

Sources of variation	*df*	Variance	F_1	F_2
Between tasks	3	87.36	40.26**	45.98**
Between groups	15	7.63	3.52**	4.02**
Between initial information	1	26.91	12.40**	14.16**
Between Task Sequences	3	5.79	2.66*	3.05*
Initial Information × Task Sequences	3	1.02	—	—
Remaining group variation	8	8.39	3.87**	4.42**
Between information	1	1.13	—	—
Groups × Information interaction	15	4.30	1.98*	2.26**
Information × Initial information	1	8.62	3.97*	4.54*
Information × Task Sequences	3	5.57	2.57	2.93*
Information × Initial Information × Task Sequences	3	9.79	4.51**	5.15**
Information × Remaining Group Variation	8	1.22	—	—
Groups × Tasks Interaction	45	.93	—	
Information × Tasks Interaction	3	.93	—	
Residual (for F_1)	173	2.17		
Combined residual (for F_2)	221	1.90		
Total	255			

* Significant at .05 level.
** Significant at .01 level.

ous interaction and many proposals of possible plays.

b. Group differences. Group differences, which were significant in their effects on the attempted leadership scores, were made more specific in the detailed analyses, involving the subpartitions of the group differences. The eight groups that began the series of four tasks with relevant information on the first task had an average attempted leadership score of 3.91 as compared with 3.27 for the eight groups that began without relevant information on the first task. It appears that initial group experiences associated with the possession of relevant information during the first task established a set toward attempting to lead.

The attempted leadership score of the group also was dependent upon the sequence in which the four tasks were presented. When the sequence began with the Discussion task (mean = 3.15), the groups attempted fewer leadership acts than when the sequence began with the Assembly task (mean = 3.72), the Strategy task (mean = 3.64), or the Construction task (mean = 3.84). The difference between the sequence headed by the Construction task and the sequence beginning with the Discussion task is significant at the .01 level. These two tasks differed in the degree to which they facilitated social interaction. While the Discussion task required some group discussion, unacquainted group members showed a strong tendency to work as individuals with the factual information supplied them. On the other hand, the Construction task required coordinated use of materials and permitted active participation by all members throughout the 15 minute interval.

c. Interaction effects. Statistically significant results bearing on the information variable are observed in the interactions involving group differences and the information variable. The interaction of Information × Initial Information is significant beyond the .05 level. Inspection of average attempted leadership scores for the four combinations involved in this interaction suggests that the type of information possessed on the initial task is particularly important in its influence on the number of leadership acts subsequently attempted under the condition of possessions of nonrelevant information. The average attempted leadership scores for the four conditions involved in this interaction are as follows:

1. For the subjects who had nonrelevant information on Trial 1, but relevant information on some other trials, the average of attempted leadership scores on trials on which they had relevant information was 3.52 ($N = 64$). The subjects and trials included in this average are as follows (refer to Table 1): Subject 1, Trials 2, 3 and 4; Subject 2, Trials 2 and 4; Subject 3, Trials 2 and 4; Subject 4, Trial 2; Subject 9, Trials 2, 3 and 4, etc.

2. For subjects who had relevant information on Trial 1, or on Trial 1 and on some other trials, the average of the attempted leadership scores on all trials on which they had relevant information was 3.80 ($N = 64$). The subjects and trials included in this average are as follows (refer to Table 1): Subject 5, Trials 1, 3 and 4; Subject 6, Trials 1, and 4; Subject 7, Trials 1 and 4; Subject 8, Trial 1; Subject 13, Trials 1, 3, and 4, etc.

3. For subjects who had nonrelevant information on Trial 1, or on Trial 1 and some other trials, the average attempted leadership scores on all trials on which they had nonrelevant information was 3.02 ($N = 64$). The subjects and

TABLE 6

Average Attempted Leadership Scores for Different Task Sequences

Task sequence	Average attempted leadership scores	
	When possessing relevant information	When possessing nonrelevant information
Assembly first	4.00	3.44
Strategy first	3.69	3.59
Construction first	4.13	3.56
Discussion first	2.81	3.50

trials included in this average are as follows (refer to Table 1): Subject 1, Trial 1; Subject 2, Trials 1 and 3; Subject 3, Trials 1 and 3; Subject 4, Trials 1, 3 and 4; Subject 9, Trial 1, etc.

4. For subjects who had relevant information on Trial 1, but nonrelevant information on some other trials, the average of attempted leadership scores on trials on which they had nonrelevant information was 4.03 ($N = 64$). The subjects and trials included in this average are as follows (refer to Table 1): Subject 5, Trial 2; Subject 6, Trials 2 and 3; Subject 7, Trials 2 and 3; Subject 8, Trials 2, 3, and 4; Subject 13, Trial 2, etc.

The difference between the average attempted leadership scores for conditions 3 and 4 is significant beyond the .01 level. It appears that initial possession of *nonrelevant* information depresses the general level of attempted leadership scores under the condition of having nonrelevant information. On the other hand, initial experience with *relevant* information tends to stimulate attempts to lead even on later trials, when the subjects do not possess relevant information.

The interaction of Information × Task Sequences is significant at the .05 level. This interaction is attributable largely to the fact that the expected relation between possessing information and attempting acts of leadership is found to reverse itself in the case of groups that began the sequence with the Discussion task. The data on which this interpretation rests are given in Table 6.

For those groups that began with the Discussion task, the average attempted leadership score for those possessing relevant information was found to be 2.81, as compared with a mean of 3.50 for those possessing nonrelevant information. It will be recalled that possession of information judged to be relevant was shown to be negatively related to performance on the Discussion task (see Section II, C, 3). Perhaps group members who have initial experiences with information that appears to be relevant, but actually may be confusing rather than helpful, come to distrust information that apparently is relevant to subsequent tasks.

TABLE 7

Average Attempted Leadership Scores Under Sixteen Experimental Conditions

Task sequence	Relevant information		Nonrelevant information	
	Initial relevant information	No initial relevant information	Initial relevant information	No initial relevant information
Assembly first	3.62	4.38	4.31	2.56
Construction first	4.44	2.94	3.35	3.44
Strategy first	4.25	4.00	3.81	3.31
Discussion first	2.87	2.75	4.25	2.75

The second-order interaction, Information × Task Sequence × Initial Information is significant at the .01 level. Second-order interaction effects are somewhat complex, but it seems worth while to examine this interaction in detail. Table 7 presents the average attempted leadership scores involved in this second-order interaction.

Apparently, the condition most conducive to attempting leadership acts had these two characteristics: (*a*) the group members possessed relevant information during their initial trial, and (*b*) they undertook the tasks in a sequence beginning with the Construction task. The least motivating condition for attempts at leadership seems to have been set by the possession of nonrelevant information, when the group began its initial task without relevant information, in a sequence beginning with the Assembly task.

A number of irregularities appear in Table 7. Note that for the task sequence beginning with the Discussion task, the experience of having relevant information on this initial task produced a larger number of attempted leadership acts under the condition of possessing nonrelevant information than under the condition of possessing relevant information. Again, this may be attributed to the inverse relation between the judged relevance of the information and task performance.

Note, also, under the condition of possessing relevant information, the case of the task sequence beginning with the Assembly task. In this instance initial possession of nonrelevant information has an effect opposite to that shown in the other task sequences. This may mean that inability to proceed with the task of assembling electronic components in the absence of clearly necessary information, produces a less depressing effect upon the group than similar experiences with more familiar tasks in which expectations of success are greater.

In general, *the results of the analysis show no simple and direct effect of information upon the attempted leadership scores.* However, the information conditions do become involved in the interaction effects associated with other experimental conditions. *Task-relevant information as supplied in this experiment seemed to function as a catalyst in the leadership process, rather than as a principal agent.*

TABLE 8

Analysis of the Variation in Attempted Leadership Scores Associated with Group Differences, Task Differences, and Trial Differences

Sources of variation	*df*	Variance	*F*
Between groups	15	7.63	3.76**
Between tasks	3	87.36	43.03**
Between trials	3	3.54	1.74
Residual	234	2.03	
Total	255		

** Significant at .01 level.

d. Trial differences. Since in the previous analysis Information was not found to be significant as a main source of variation, it was possible to disregard this experimental variable and test the significance of Trial differences on the attempted leadership scores. Group differences, Task differences, and Trial differences are available for this analysis, which is presented in Table 8.

We are interested in only the *F* ratio for between trials, which fails to meet the .05 level of significance. It appears, therefore, that there is no significant difference in the attempted leadership scores associated with whether the groups were working on their first, second, third, or fourth task.

e. Individual differences. Since there are no differences in number of attempted leadership acts from trial to trial, we may well consider the consistency of this measure for the individual. However, each trial represented a new task for the individual, and it has been demonstrated that there are significant differences in number of attempted leadership acts as a function of Task differences. Accordingly, in order to eliminate this effect, which would tend to attenuate any measure of consistency, the scores for each task were standardized by converting them to McCall *T* scores. As a measure of the consistency with which individuals attempted to lead, an intraclass correlation coefficient was computed. The obtained intraclass correlation of .45 indicates that a consistency in attempts to lead may characterize the individual. However, a correlation of .45 is relatively small, suggesting that attempted leadership is not primarily a function of individual differences, but presumably is also related to situational characteristics. It should be noted, too, that the group composition remained intact through the four tasks, which condition might be expected to maximize the tendency for an individual to retain, throughout the entire session, the status he acquired as a leader on an early trial.

2. *Further Analysis of the Information Variable*

By removing the effects of these individual differences in general disposition to attempt to lead, we were able to obtain a more sensitive test of the situational variable that is of primary interest —possession of information—in that the effects of possession of information may have been obscured by individual differences. It has been shown previously that trials have little effect on the variance of the scores, and the effects of differences among the four tasks have been removed by standardizing the scores.

a. Removal of individual differences. It was possible, within the basic experimental design, to remove the effects of individual differences in disposition to attempt to lead. Since all individuals were subjected to the same conditions with respect to possession or nonpossession of relevant information on Trials 1 and 2, it is assumed that differences which appear in these trials, other than those due to tasks and trials, are due to individual differences in "natural" predisposition to lead.

After attempted leadership scores (standardized for task differences) on Trials 1 and 2 were correlated with scores on Trials 3 and 4, regression equations were written to predict separately scores on Trials 3 and 4. These predicted scores represent that portion of the obtained scores that can be accounted for in terms of performance on Trials 1 and 2.

The multiple correlations of Trials 1 and 2 with Trials 3 and 4 were .52 ($R_{3.12}$) and .68 ($R_{4.12}$) respectively. The corresponding raw score regression equations are:

$$X_3 = .1776X_1 + .3497X_2 + 24.60$$

and

$$X_4 = -.0726X_1 + .7074X_2 + 18.97.$$

From these equations, predicted scores for Trials 3 and 4 were generated and subtracted from the obtained scores. These difference scores are independent of both task differences and individual differences.

b. Analysis of difference scores. In order to determine the effects of possession or nonpossession of relevant information on these scores, *t* tests involving the following comparisons were carried out:

1. The mean attempted leadership difference score of individuals possessing task-relevant information was compared with the mean attempted leadership difference score of individuals possessing no task-relevant information.

2. In the case of the Construction and Assembly tasks, the mean attempted leadership difference score of those in-

dividuals who had scored 0, 1, or 2 on the reading comprehension test, was compared with the mean attempted leadership difference score of those individuals who had scored 3, 4, or 5 on the test.

The first of these comparisons is regarded as the most general test of the effect of possession of relevant information on attempted leadership. The *t* ratio resulting from this comparison had a value of .92, which is not significant. Since this test did not reveal a significant effect, the second and more restricted comparison was carried out as a final means of determining whether possession of relevant information might be shown to affect attempts to lead.

It was reasoned that the absence of any effects due to possession of information might be due to an improper definition of "relevant" information. It was conceivable that the subjects had brought with them to the experiment information that was relevant to the tasks and that was not appreciably supplemented by the information made available during the course of the experiment. In that event, scores on the reading comprehension test would be a better indication of whether or not the subjects did, in fact, possess task-relevant information. Accordingly, for the purposes of the second *t* test referred to above, possession of task-relevant information was defined in terms of performance on the reading comprehension test. Those who scored 3, 4, or 5 were classified as possessing information; those who scored 0, 1, and 2 were classified as not possessing information. A further restriction was imposed by testing only those attempted leadership scores observed on the Construction and Assembly tasks, since these were the only tasks for which differences in performance were found. The purpose of this restriction was to provide as critical a test as possible by limiting the test to those tasks for which differences in performance had been observed. Despite these considerations, the null hypothesis was accepted ($t = .53$).

In brief, it appears that *possession of task-relevant information, in and of itself, does not affect attempted leadership scores.* This is true even when we define possession of information in terms of reading comprehension test performance rather than in terms of the experimental manipulation of the information booklets.

IV. DISCUSSION OF RESULTS

The principal hypothesis of this study was not directly supported by the results of the experiment. Possession of information judged to be relevant to the solution of the groups' tasks was not shown to be a factor directly associated with greater frequency of attempts to lead. We have no evidence, therefore, to conclude that possession of task-relevant information per se is a major determinant of motivation for leadership.

Nevertheless, the results of the experiment do point to important special conditions under which relevant information is associated with greater frequencies of attempts to lead. In the following discussion of the results we shall be concerned with four major issues: (*a*) the alternative uses to which task-relevant information may be put by group members; (*b*) the factors determining the relevance of information for group problem solution; (*c*) task characteristics associated with the frequency of attempts

Bureau Ednl. Psy. Research
DAVID HARE TRAINING COLLEGE
Dated.........
Acc. No.........

to lead; and (*d*) changes in group behavior relative to the experiences of group members with information having varying task relevance.

It is believed that the discussion of these four issues will point to some of the reasons why the simple possession of task-relevant information cannot be used as a general predictor of attempted leadership.

A. ALTERNATIVE USES OF TASK-RELEVANT INFORMATION

The possession of task-relevant information may motivate an individual to assume the role of either "leader" or "resource person." Both of these roles were observed frequently among the subjects of the experiment. Both roles apparently were acceptable to the subjects. The fact that just prior to their problem-solving exercise they were exposed to and tested on information having some bearing on their tasks, may have suggested the potential value of the "resource person" role. In retrospect, it appears that failure to control feedback during the sessions may have been a major factor in facilitating a redistribution of information by a voluntary resource person. In the experimental instructions, no emphasis was placed on leadership or organization of group effort, and rewards were offered for group performance rather than for successful leadership. It has been noted, also, that the groups were composed of relatively homogeneous and previously unacquainted individuals. These conditions, too, may have helped to define the particular situation as one in which there was no special incentive to try to assume a leader role in preference to that of a resource person.

Another deviation from the direct use of relevant information to lead appeared when one group member engaged in extracting relevant information from another. On occasion, a dominant individual, or one who seemed previously to have gained some status as a leader, would ask for and receive task-relevant information from another group member and then proceed to utilize this borrowed information in attempting to lead. In one sense, the group member who supplied the information was a "resource person." But he differed from those who spontaneously volunteered their information to the group in that he did not seem consistently to see himself in the "information giver" role. Examination of transcripts of the sessions makes it apparent that the relationship between possession of information and member function in the group process can be highly complex.

B. FACTORS DETERMINING THE RELEVANCE OF INFORMATION

The effectiveness of the experiment was compromised in part by results that raise serious questions about the degree to which we were able to manipulate the relevance of information variable. We were able to achieve high agreement among judges who appraised each of the eight sets of information, with respect to which set within a given pair was more relevant to the task described. In addition, it was demonstrated that those individuals who had been provided with information judged to be relevant achieved significantly higher scores on both the comprehension and the retention tests.

In the more critical test, however, of whether the information was relevant in the sense that it contributed to better group task performance, significant results were obtained for only two of the tasks. In fact, in the case of the Discussion task, there was a slight tendency for the groups given the nonrelevant information to do better than those possessing the so-called relevant information. Although this discrepancy is a minor trend in a small sample of observations and might be dismissed on this basis alone, there may be something to be gained by re-examining the concept of relevance of information.

Perhaps the concern with the relation between possession of information and actual group task performance is misplaced. We were interested primarily in the manipulation of information in order to examine the possibility of associated variation in frequency of attempting to lead. Presumably, any such association would be mediated through the perception by the group members of the relevance of their own information to the solution of the group's problem. If this were the case, then it might be reasoned that relevant-appearing information would be more likely to stimulate attempts to lead than information that was not obviously relevant, even if it proved to be useful in practice. A fault in this reasoning, however, may lie in the contrary effects of accumulated experience with so-called relevant information that actually was found to be of no use. Unfortunately, the information supplied in the case of the Discussion task seemed to be largely of this type. Many facts about voting behavior were supplied by the booklet of relevant information. Only three or four of these facts could be employed in making a rational estimate of the number of votes cast in Oregon for the Republican candidate for president in 1940. The other facts, although relevant to the prediction of voting behavior in general, had nothing to do specifically with Oregon, or the year 1940, or Republican voting. Hence, these items may have added only confusion.

In the cases of the two tasks, Assembly and Construction, for which a significant relation was found between possession of the relevant information and task performance, a greater proportion of the information was directly usable. A good performance on the Construction task was virtually impossible without knowledge of the manufacturer's diagram against which the group product was scored. Over three-quarters of the specific items of information supplied for the Assembly task were needed for a perfect performance. Although for both of these tasks some surplus information was supplied, there appeared to be a relatively small amount that could be regarded as misleading.

Another aspect of the problem of the manipulation, or distribution, of relevant information related to the novelty for the group members of the various tasks. The screening of subjects did eliminate from the experiment individuals who had had specific experience with obviously similar tasks. Nevertheless, the four tasks varied greatly in terms of the degree to which "everyday common sense" might appear to afford an adequate basis for solution of the problem. The Strategy and Discussion problems seemed most frequently to be approached with efforts to use previously acquired knowledge and experience. The Assembly task probably was least familiar to most of the subjects, although occasionally some incidental knowledge (acquired through tinkering with the family radio) was brought to bear on this problem.

Some of the findings of the analysis involving task and information sequences and their interactions may be more understandable in view of the special characteristics of the four tasks and the complex relations among tasks, information, and attempts to lead.

C. Task Characteristics Associated with the Frequency of Attempting to Lead

The largest single source of variation among the attempted leadership scores was task differences (See Section III, B, 1, 2). In other words, special characteristics of the tasks had much to do with how frequently group members tried to lead. By far the largest frequency of attempts to lead was observed when the groups were working on the Strategy task, which required the four group members to work as a team in competition with the experimenter in a modified game of peggity. Before each move, the four group members were required to decide upon the deployment of their four pegs. This rule, in turn, meant that the group had to reach consensus with each move. Therefore, the interaction among the group members was lively and characterized by a large number of attempts to lead. Polite reticence quickly disappeared in the face of direct, positive recommendations of hastily conceived plans that were just as frequently rejected by other members' equally hearty endorsement of some other plans. The pressure of time and the recurring necessity of arriving at consensus elevated the frequency of attempted leadership acts above that observed for any other of the four tasks.

In contrast to the rapid interaction and full participation that were noted in the Strategy problem, there was relatively little interaction in the Discussion problem. The typical sequence of events included a short general discussion at the beginning of the fifteen minute period, followed by an interval during which individuals worked alone on various computations, then engaged in some comparing of results, and, finally, near the close of the allotted time, a short period of active participation by all. At this point, agreement was reached upon a final estimate, the only time at which consensus was required.

The Construction task differed from the others primarily in the kind of atmosphere created for the group. Work on this task involved much gross muscular activity with the members helping each other to manipulate the materials. Since it was quite easy for all group members to work simultaneously on different parts of the task, participation started and remained at a high level throughout the period. The type of activity required and the amount of participation stimulated both served to make this task an excellent "ice-breaker" for previously unacquainted individuals. These characteristics of the task may have accounted for the fact that more leadership acts were attempted by members of groups who began their sequence of four tasks with the Construction problem.

The Assembly task seemed to be extremely frustrating for group members who had not received relevant information. Occasionally, a rather prolonged period of silence was noted at the beginning of such a session. Participation was moderately high, however, and some agreement was required in order that a particular component could be placed in the circuit. Such informal observations have made it increasingly obvious to us that much more research is needed on task characteristics per se as predictors of leadership behavior in groups.

D. Changes in Group Behavior

In the analysis of the experimental data we noted the significant effects on attempted leadership scores of two conditions: (*a*) whether the group began its initial task with or without relevant information, and (*b*) the sequence in which the four tasks were presented. Although this is not the only difference between the sequences, this sequence effect may be regarded as a function of the initial task in the sequence. If we do so regard the sequence and also note the effect of information on the initial trial, we can infer that initial group experiences may have modified subsequent attempts to lead.

Following the above interpretation of these effects, we find that the possession of relevant information on the initial trial does tend to influence attempts to lead on subsequent trials. For the groups that had no relevant information on Trial 1, the mean attempted leadership score for Trials 2 through 4 is 3.38, to be compared with a mean of 4.03 for groups with relevant information on the first trials. Moreover, those group members who began their tasks with either the Construction or Assembly problem had an average attempted leadership score of 4.02 for the subsequent three tasks, to be compared with a mean score of 3.39 for those members whose initial experience was with the Discussion or Strategy tasks. It will be recalled that the latter two tasks were those for which the supplied information was of less clear-cut relevance to task performance. It seems evident, therefore, that the possession of useful information on the initial task of the series tended to raise the succeeding frequency of attempts to lead. Even in a brief period of fifteen minutes some group norm or set that affected this type of group member behavior appears to have been established.

E. Implications

The experiment was designed to test a hypothesis derived from central theorems of a theory of leadership and group behavior (23, Appendix A).

The experimental hypothesis stated in the null form was as follows: In a mutual problem-solving situation the frequency of attempts to lead by those individuals possessing task-relevant information is not significantly different from the frequency of leadership acts attempted by individuals possessing nonrelevant information. At the time the hypothesis was formulated, we assumed that initial possession of task-relevant information would insure that the individual would be able to see a clear relation between this relevant information and the solution of the experimental tasks (mutual problems). Therefore, his estimate of the probability of the act's leading to mutual problem solution would be higher than if he did not possess relevant information. The hypothesis still appears to us to be related to the theorem, but it now seems likely that control was not maintained over certain other variables stipulated by the theory as also being related to attempted leadership.

For example, we attempted in the initial assembly of the groups, to control variables related to individual member differences (i.e., . . . 2, his estimate of the probable effect of the act on the potential of the group for need-disposition satisfaction, 3, the degree of dissatisfaction he feels with the mutual problem, and 4, the strength of his relevant need dispositions). These controls were provided through offering uniform incentives for excellence of performance, selection of group members from a relatively homogeneous sample (i.e., male introductory psychology students, who passed our screening questions), an insistence upon members' being unacquainted before the experiment, and the use of standard experimental procedures. Although some variance may have been introduced by the inadequacies of these controls, we have no reason to believe that these were the major problems.

The results of the experiment and our informal observations both point to changes occurring in the groups themselves during the experimental sessions as the chief uncontrolled source of variation. Thus, the controls introduced at the outset do not appear to have been maintained throughout the experiment. During the initial period of their work together, the subjects rapidly became something other than unacquainted persons, developed more or less esteem in the group, learned that certain of the facts they possessed were useful and that others were not, etc. We have noted evidence suggesting that the status of the information variable may have changed during the experiment. After a brief period of group interaction, the relation of initial possession of relevant information to the individual's "probability estimate" becomes increasingly complicated.

A group member's estimate of the probability that his contemplated leadership act will result in problem solution appears to undergo revision influenced by his experience in attempting to use his "relevant" information. Such shifting was not controlled. Instead, an implicit assumption had been made that no shift would occur. In effect, therefore, we had chosen the alternative of maintaining an indirect rather than a simple and direct relation between the hypothesis tested and the proposition stated in the theory. This relationship is indirect, or qualified, in the sense that the hypothesis, with reference to a number of variables theoretically specified as pertinent to attempted leadership, has assumed that such variables would be assigned the status of "other things being equal." The experiment appears to have provided a test of the hypothesis as stated, but the evidence at hand does not seem to warrant revision of the theory itself. The experience with this attempt to test the propositions of the theory suggests a number of variables that must be taken into account in small group experiments—particularly serial, time, and feedback effects. We can point to the dangers involved in assuming that small groups of laboratory subjects will remain without a "tradition" for even very short periods of time. The research worker concerned with investigating the behavior of small groups cannot assume that, once he has manipulated his independent variable, his manipulation will remain in effect throughout his experiment.

V. SUMMARY AND CONCLUSIONS

In order to investigate the relationship between possession of task-relevant information and attempted leadership behavior, sixteen groups of four male subjects each were given four group problems—Assembly, Strategy, Construction, and Discussion tasks—to solve. Task-relevant information was supplied to various members of each group on each task during a training period preceding the groups' attempts to solve the problems. Immediately following the training period, and again after the completion of the experimental session, each group was tested to determine the extent to which the group members retained the task-relevant information supplied them. During the experimental session three trained observers located behind one-way vision mirrors tallied the frequency of occurrence of attempted leadership acts. When available, a "naive" observer rated members of the groups on "amount of leadership."

Interobserver reliability of the observations of attempted leadership acts was estimated by means of the intraclass correlation coefficient (r'). The intraclass correlation among the three trained observers was .83. For the three possible pairs of observers, the intraclass correlation coefficients were .80, .81, and .88. These values indicate that trained observers are able to recognize and record attempted leadership acts with a high degree of consistency.

The ratings of the naive observers on amount of leadership were correlated with the number of attempted leadership acts reported by the three trained observers. This product-moment correlation coefficient was .47. The magnitude of this correlation indicates that, while there is some communality between the concept of the attempted leadership act and the layman's notion of leadership, the two concepts are largely independent.

As a check on the meaningfulness of the experimental classification of information as more or less task-relevant, six individuals not associated with the research group were asked to judge the information given the subjects. For three tasks, there was unanimous agreement among the judges as to how the information should be classified. The probability that such agreement would occur by chance alone is less than one in a thousand. On the Assembly task five of the six judges agreed with the experimental classification of the information. This level of agreement would be expected to occur by chance less than four times in one hundred.

As another means of assessing the distribution of information, the subjects were tested on the task-relevant material both immediately after the training period and also following the completion of their tasks. In each instance, it was found that subjects who had had access to more task-relevant information scored significantly higher on the tests than those who had been presented with nonrelevant information. Insofar as performance on multiple-choice item tests is concerned, the distribution of information among the experimental subjects was in accordance with the experimental design.

Task performance was not so clearly related to the information variable. On two of the tasks (Assembly and Construction) individuals who possessed relevant information scored significantly higher than those who did not possess such information. On the other two tasks, Strategy and Discussion, there were no significant differences between the mean performance scores of the two classes of subjects. On the latter task, however, there was a slight trend in a reversed direction, i.e., a tendency for those subjects who possessed nonrelevant task information to receive superior task performance scores.

The principal hypothesis under investigation in this experiment is that in a group problem-solving situation individuals who possess task-relevant information will attempt a greater number of leadership acts than individuals who possess no task-relevant information. The results of the experiment do not support this hypothesis. However, certain interaction effects involving information were found to be significantly related to attempts to lead.

Three interactions—two first-order and one second-order—involving the information variable were found to be significant. These were (*a*) Information × Initial Information, (*b*) Information × Task Sequence, and (*c*) Information × Initial Information × Task Sequence. The fact that the Information × Initial Information interaction was significant indicates that possessing (or not possessing) information has an effect upon attempted leadership scores that depends upon whether or not the subjects were supplied with relevant information on their *initial* task in the sequence of four tasks. The largest average attempted leadership score was obtained in the case of subjects who had begun their sequence of four tasks *with* relevant information, but who worked *without* relevant information on one or more of the succeeding task. (See Section III,B,1,c for a more complete statement of this interaction effect.)

The Information × Task Sequence interaction was found to be significant. This is due to the departure from the expected attempted leadership scores when groups began their experimental sessions with the Discussion task. Under this condition, attempted leadership scores were higher for those individuals who did not possess relevant information. This reversal appears to be related to the similar reversal observed in the case of performance scores on this task. Apparently the information that was experimentally classified as "relevant" to the Discussion task served to confuse the subjects. Accordingly, initial experience with seemingly relevant information that actually was of little use to the group served to depress the ensuing level of attempted leadership acts for such groups.

No clear-cut interpretation of the complex effects present in the second-order interaction of Information × Initial Information × Task Sequence is readily at hand. This interaction, taken in conjunction with the preceding interaction effects, indicates that while information has no simple and direct effects upon attempted leadership, it is related to other situational variables.

The principal conclusions drawn from the results of this experiment may be summarized as follows:

1. Under the conditions of the ex-

periment, possession of task-relevant information was shown to have a complex rather than a simple relation to the frequency of attempts to lead.

2. The number of leadership acts attempted by group members during the fifteen minutes allotted for the solution of each of four different tasks differed significantly among the tasks.

3. Individuals displayed a significant, but moderate, consistency in their disposition to attempt leadership.

4. Observers can be trained to make reliable observations of attempted leadership acts.

REFERENCES

1. Ansbacher, H. L. The history of the leaderless group discussion technique. *Psychol. Bull.*, 1951, 48, 383-390.
2. Bales, R. F. *Interaction process analysis.* Cambridge: Addison-Wesley, 1950.
3. Bass, B. M. An analysis of the leaderless group discussion. *J. appl. Psychol.*, 1949, 33, 527-533.
4. Bass, B. M., Wurster, C. R., Doll, Paddy Ann, & Clair, D. G. Situational and personality factors in leadership among sorority women. *Psychol. Monogr.*, 1953, 67, No. 16 (Whole No. 366).
5. Bell, G. B., & French, R. L. Consistency of individual leadership position in small groups of varying membership. *J. abnorm. soc. Psychol.*, 1950, 45, 764-767.
6. Caldwell, O. W. Some factors in training for leadership. *Nat. Asso. Sec. Sch. Prin., Fourth Yearb.*, 1920. Pp. 2-13.
7. Carter, L., Haythorn, W., Meirowitz, B., & Lanzetta, J. A note on a new technique of interaction recording. *J. abnorm. soc. Psychol.*, 1951, 46, 258-260.
8. Carter, L., Haythorn, W., Meirowitz, B., & Lanzetta, J. The relation of categorizations and ratings in the observation of group behavior. *Hum. Rel.*, 1951, 4, 239-254.
9. Carter, L., Haythorn, W., Shriver, B., & Lanzetta, J. The behavior of leaders and other group members. *J. abnorm. soc. Psychol.*, 1951, 46, 4, 589-595.
10. Deutsch, M. An experimental study of the effects of cooperation and competition upon group process. *Hum. Relat.*, 1949, 2, 199-231.
11. Deutsch, M. A theory of cooperation and competition. *Hum. Relat.*, 1949, 2, 129-152.
12. Deutsch, M. *A study of conditions affecting cooperation.* Annu. Rep., No. 2, Contract No. NR-285 (10), Human Relations Branch, Office of Naval Research, 1954 (mimeographed).
13. Fisher, R. A. *Statistical methods for research workers.* London: Oliver and Boyd, 1938.
14. Gibb, C. A. The principles and traits of leadership. *J. abnorm. soc. Psychol.*, 1947, 42, 267-284.
15. Gibb, C. A. The research background of an interactional theory of leadership. *Australian J. Psychol.*, 1950, 28, 10-42.
16. Gibb, C. A. An experimental approach to the study of leadership. *Occup. Psychol.*, 1951, 25, 233-248.
17. Guetzkow, H. Unitizing and categorizing problems in coding quantitative data. *J. clin. Psychol.*, 1950, 6, 47-58.
18. Guetzkow, H. (Ed.) *Groups, leadership and men.* Pittsburgh: Carnegie Press, 1951.
19. Halpin, A. W., & Winer, B. J. *The leadership behavior of the airplane commander.* Washington, D.C.: Human Resources Research Laboratories, Dept. of the Air Force, 1952.
20. Heise, G. A., & Miller, G. A. Problem solving by small groups using various communication nets. *J. abnorm. soc. Psychol.*, 1951, 46, 327-336.
21. Hemphill, J. K. Situational factors in leadership. *Ohio State Univer. Bur. Educ. Res. Monogr.*, 1949. No. 32.
22. Hemphill, J. K., & Coons, A. E. *Leader behavior description.* Columbus: Personnel Research Board, The Ohio State Univer., 1950.
23. Hemphill, J. K., Pepinsky, Pauline N., Shevitz, R. N., Jaynes, W. E., & Christner, Charlotte A. *Leadership acts: 1. An investigation of the relation between task relevant information and attempts to lead.* Columbus: Ohio State Univer. Res. Found., 1954.
24. Hemphill, J. K., & Westie, C. M. The measurement of group dimensions. *J. Psychol.*, 1950, 27, 325-342.
25. Heyns, R. W., & Zander, A. F. Observation of group behavior. In L. Festinger and D. Katz (Eds.), *Research methods in the behavioral sciences.* New York: Dryden, 1953. Pp. 381-417.
26. Jenkins, W. O. A review of leadership studies. *Psychol. Bull.*, 1947, 44, 54-79.
27. Johnson, P. O. *Statistical methods in re-*

search. New York: Prentice-Hall, 1949.

28. Katz, D. Social psychology and group processes. *Annu. Rev. Psychol.*, 1951, **2**, 137-172.
29. Leavitt, H. J. Some effects of certain communication patterns on group performance. *J. abnorm. soc. Psychol.*, 1951, **46**, 38-50.
30. Newcomb, T. M. Motivation in social behavior. In J. S. Brown, *et al., Current theory and research in motivation: a symposium*. Lincoln: Univer. of Nebraska Press, 1953. Pp. 139-161.
31. Roseborough, Mary E. Experimental studies of small groups. *Psychol. Bull.*, 1953, **50**, 275-303.
32. Shartle, C. L., & Stogdill, R. M. Methods of determining patterns of leadership behavior in relation to organizational structure and objectives. *J. appl. Psychol.*, 1948, **32**, 286-291.
33. Smith, M. B. Social psychology and group processes. *Annu. Rev. Psychol.*, 1952, **3**, 175-204.
34. Stogdill, R. M. Personal factors associated with leadership: a survey of the literature. *J. Psychol.*, 1948, **25**, 35-71.
35. Stogdill, R. M. Leadership, membership, and organization. *Psychol. Bull.*, 1950, **47**, 1-14.

(Accepted for publication October 16, 1955)

Vol. 70, No. 8 Whole No. 415, 1956

Psychological Monographs: General and Applied

Loci of Interference in a Complex Maze[1]

Albert H. Smith

Queen's University, Kingston, Ontario

There have been relatively few studies of interference (retroactive inhibition) with animal subjects. Of eight previous investigations, five yielded a relearning decrement and three, facilitation. With one exception (11) hypotheses have lacked precision and the data and analyses have been coarse, perhaps understandably since the investigators were concerned mainly with demonstrating an over-all interference effect. However, the more numerous transfer experiments with animals, reviewed by Britt (5) and Kogan (11), have raised two general issues relevant to interference which are amenable to study within the framework of the similarity concept, in terms analogous to those which have governed much of the verbal-learning research. The first is the question of the relation between negative transfer and the principles which describe animal maze performance. A few of these principles encompass the influences on the running response of local structural features of the maze, but the majority embody the view that rats form generalizations of the maze pattern which determine the nature of the running response on successive trials. With the exception cited above, attempts to relate these principles to interference have been lacking. The second issue arises from the fact that opinion is almost evenly divided on the question as to whether negative transfer is confined to the loci of changes in the second task of the transfer series (4, 8, 9, 19), or whether it spreads to neighboring unchanged elements (6, 7, 13, 18, 20). It is clear that these questions are related, to the extent that the answer to the first may provide the answer to the other; but from the point of view of method, the second, one might well be investigated independently.

This is an exploratory study with the following aims: (*a*) to produce interference in rat maze learning with acceptable generality, (*b*) to define its relationship to the response patterns identifiable in a complex motor activity, and (*c*) to determine whether or not interference is specific to task components. The three major variables in as many experiments were the similarity factor, temporal point of interpolation, and degree of interpolated practice. The maze design was complex enough to support the expectation that it would elicit such response tendencies as a gradient of elimination of errors (10), anticipatory behavior (16), situationally determined responses (1, 3, 15), and perhaps, as a function of its complexity, some sort of "abstractive" behavior (12). Such complex tasks have sometimes failed to produce interference, and the response patterns therein have typically been difficult to analyze. The ideal of hypothetical specificity clearly demands a simpler task, but this would

---

[1] This monograph is condensed from a thesis submitted in partial fulfilment of the requirements for the Ph.D. degree at the University of Toronto.

Bure[illegible] Ednl. [illegible]sy. Research
DAVID HA[illegible] [illegible]ING COLLEGE
Dated
No

have forestalled the possibility of producing interesting response tendencies associated with the complexity of the task.

I. METHOD

Subjects

The subjects were 151 albino rats matched in groups of approximately 10—half males, half females to a group. Those of the first two experiments were obtained from a farm colony. Their mean age was 82 days, with a range of 15 days, at the start of preliminary training. Those of the third experiment were bred in the departmental laboratory from animals selected from those of the first experiment. Six litters were allocated to groups by splitting the litters as far as was consistent with matching. Their mean age at the start of preliminary training was 75 days, with a range of three days. Two died and two were discarded for "freezing" during preliminary training in the first experiment. Two were discarded in the original learning session of the second experiment, and one in the same session of the third, for slow running and excessive retracing. The others remained in good health. They were weighed once a week just before feeding. The curves for mean weight gains indicate that a uniform hunger drive was maintained. It was probably stronger in the second and third experiments, since initial weight loss was greater and performance better in terms of errors and running speed. The Chow cubes obtained for these two experiments weighed about 10 per cent less than those used in the first. The percentages of individual weight loss during the first two weeks (O.L. sessions) of these experiments had zero correlation with the individual first-choice error scores.

Apparatus

The maze was an alley-type, 14-unit, multiple-T maze made of half-inch pine, with an open top covered with fine wire screen and painted flat black inside and out. The T-shaped sections were identical with inside dimensions 4 inches square, a 16-inch crosspiece, and a 6-inch stem. The entrance and goal boxes were 10 inches square, 4 inches deep. Drop-type metal doors, eight in all, were placed across the entrance and goal boxes and at the junctions of all units except the first and last two. A straight, 3-foot section of the same alley dimensions was used for preliminary training. It was decided to dispense with the desirable feature of crosspieces at the ends of blinds in order to expedite mastery, hence reduce the total period of food deprivation, and also to facilitate making four changes of pattern daily in the interpolated learning session of Experiment I. The maze rested on the floor of an experimental room separate from the vivarium, and was surrounded by a homogeneously colored screen. The room was illuminated by a 40-watt frosted lamp enclosed by a frosted shade suspended centrally over the maze.

Experimental Design

Five maze patterns were employed in the standard interference paradigm with the following series of correct turns. The O.L. and R.L. task in all experiments was Pattern A, the double alternation scheme, LLRRLLRRLLRRLL. The I.L. tasks in Experiment I were: Pattern B, the mirror image of Pattern A, RRLLRRLLRRLLRR; Pattern C, a modified single alternation scheme, LRLRLRLLRLRLRL; Pattern D, a single-double alternation scheme, LRLLRLRRLRLLRL; and Pattern E, a more complex arrangement, RLLRLLR-RLLRLLR. Pattern C was the single interpolated task in Experiments II and III since it produced most interference in Experiment I. The experimental groups are designated by the letters of their respective interpolated patterns, so that "group" and "pattern" are generally interchangeable.

In the O.L. sessions, all animals were given criterion training on Pattern A and were then separated into matched groups on the basis of their first-choice errors. The experimental groups were given the interpolated tasks while the controls were left in their cages but were otherwise treated the same, including feeding. No attempt was made to give the controls equivalent exercise. All then relearned Pattern A to the criterion. The I.L. sessions were as follows: in Experiment I, beginning the day after the last rat finished O.L., the four experimental groups received criterion training on Patterns B, C, D, and E, one group to a pattern. In Experiment II, the three experimental groups, C1, C2, and C3, received criterion training on Pattern C beginning, respectively, on the first, eleventh, and twenty-first day after O.L. In Experiment III, the four experimental groups, C1-C4, received 5, 10, 15, and 20 trials, respectively, on Pattern C, beginning the day after O.L. In Experiment I, each experimental group began R.L. as a group on the day after it finished I.L. Control Group 1 began R.L. with the first experimental group, and Control Group 2 with the last experimental group, to finish I.L. In Experiments II and III, all groups began R.L. on the day after the last rat finished I.L. In Experiment II, one rat failed to reach the criterion of mastery in I.L. and one failed in R.L.

The first experiment represents the major project. The other two were conducted with the hope of confirming an interference effect, in order to obtain comparative data on the effects of interpolated periods of no practice of the magnitude encountered in Experiment I, and to make it possible to specify more exactly the nature of the O.L. task and that of the I.L. task which produced most interference in this experiment. Comparisons among the results of the three experiments should be made in the light of the unexpected fact that the experiments extended over 8, 10, and 5 weeks, respectively. In Experiment II, Group C3 required longer for I.L. than was predicted, and in Experiment III the homebred rats turned in a superior performance.

The I.L. conditions of Experiment I were expected to differentiate the groups in terms of trials to criterion, but there was no basis at this stage for predicting how much. Rather than delay the start of R.L. until the last rat finished I.L. and possibly incur a relearning decrement, it was decided to start each group on R.L. the day after the last rat in the group finished I.L. and to test the alternative procedure in the other experiments. Also, two control groups were used, which began R.L., respectively, with the earliest and latest experimental groups to finish I.L., Groups C and B, respectively. Groups C, D, and E reached the criterion within three days of each other, hence their scores should be compared with those of Control Group 1. Group B required 10 more I.L. trials than Group C, hence its performance is referable to that of Control Group 2. Further reference to the time factor appears in the section on results.

In Experiment II, the delay periods were imposed immediately after O.L. because the use of criterion training made it most convenient to specify them at this stage. Further, the experimental groups of Experiment I finished I.L. over an interval of 10 days and it was expected that, in Experiment II, the "inactivity" period of at least one of the groups would be about 10 days, thus providing an index of the interference effect of delays of this order. Also, Group C2 in this experiment had an imposed delay of 10 days before beginning I.L. In Experiment III, interpolation directly after O.L. is the most convenient arrangement for imposing variations in the amount of practice. This experiment was also expected to provide an indication of the effects of imposing different periods of "inactivity" between the end of I.L. and the beginning of R.L.

Previous research provided few criteria for selecting from the very large number of patterns of turns possible with this maze those likely to produce marked interference. The I.L. tasks must be similar to the O.L. task and at least three degrees of similarity are required. A maze with identical reversible sections permits such variations without changing the serial positions of the units. The linear maze comes closest to producing an uncomplicated goal gradient when this is the primary interest. In most of the present patterns the line of progress is roughly linear and the series of turns, including numbers and positions of reversals, are the same in the first and second halves. The double alternation task, Pattern A, was chosen arbitrarily for O.L. In terms of number of reversals, its mirror image, Pattern B, is maximally dissimilar with 14; but if rats can generalize complex response patterns, this task could be expected to become functionally more like Pattern A. Patterns D and E, with eight and four reversals, respectively, represent increasing degrees of similarity to Pattern A. It also seemed desirable to interpolate a less difficult task, a single alternation pattern. However, the experimental room was too small to accommodate 14 single alternations, so two left turns were inserted at its mid-point to yield Pattern C. This, also, has 8 reversals compared to Pattern A.

Training Procedure

All training began at 5:30 P.M. and was at the rate of one trial a day. Preliminary training consisted of five trials in the straight runway. The running order in all series except the I.L. session of Experiment I was half of the males, half of the females, which was repeated, always in the same individual order. In the I.L. session of Experiment I, the rats were run by groups in the order, B, C, D, and E, males first within groups. At the end of runs, the rats were permitted to nibble the food in the goal box for a few seconds and were then fed three-grain poultry wheat for 30 minutes in separate feeding cages in the vivarium. Males consumed on the average nine grams, females eight. After this, each rat received a cube of Fox Chow (6-7½ grams) which was completely consumed. Greens were given once a week instead of Chow. The minimum period of food deprivation was 22 hours. The performance indices were: (*a*) first-choice errors, in which the rat placed at least both feet in the blind when at a choice point for the first time in the run; (*b*) retrace errors, defined as retracing any segment of the true path or a blind, each unit being regarded for this purpose as having two segments to the crosspiece and one to the stem; (*c*) seconds per run; and (*d*) number of trials to reach the criterion. Running time per unit was not recorded. A fairly stringent criterion of three successive errorless runs (absence of first-choice errors) was imposed in order to set up stable habits.

II. RESULTS

Ancillary Data

Performance curves. First-choice errors in all sessions exhibited smooth, negatively accelerated curves in contrast with the irregular curves for time scores, and especially for retracing. This is a common finding which may be interpreted as supporting the view that first-choice errors measure primarily "cognitive" functions, whereas the other indices reflect motivational and disruptive processes.

Matching. The groups were matched on the basis of first-choice errors in O.L. to yield almost identical group means in each experiment. Concurrently, close matching of mean numbers of trials and mean running speed per group were also achieved. In terms of t ratios and F ratios, none of the intergroup differences for any performance index in the O.L. sessions of the three experiments was statistically significant. The largest t ratio was .75, and most of them were close to zero. These data have been omitted from most of the subsequent tabulations.

Maze reliability. The product-moment correlations between the first-choice errors in the odd and even trials of the first 20 trials of Experiments I and II and the first eight O.L. trials of Experiment III were $.83 \pm .006$, $.95 \pm .016$, and $.74 \pm .006$, respectively. Split-half coefficients for these trials were $.82 \pm .006$, $.69 \pm .009$, and $.73 \pm .006$, respectively. All coefficients were corrected for attenuation by the Spearman-Brown formula.

Sex differences. Each group was equated for sex and the comparisons were based on mean group scores. The variance in first-choice errors due to sex was not large enough to warrant removing it from the residual term in testing for significance, since the F ratios for interference based on the pooled data were much larger than required for significance. Aside from control, however, the small but significant female superiority evident in these data is opposed to previous findings of small but significant differences favoring males. This is probably the result of greater drive since, by intention, the females received from 10 to 15 per cent less Chow than the males (smaller cubes), which may have been disproportionate to need.

The time factor. The imposition of criterion training gave rise to a problem of control of the time factor in the I.L. sessions. It was assumed to be optimal for negative transfer that stable habits of about equal strength be established in the O.L. and I.L. sessions. The following outcomes are relevant. (*a*) All groups were closely matched in terms of mean number of trials in the O.L. sessions. (*b*) In the R.L. session of Experiment I, the mean difference in first-choice errors between the two control groups was statistically insignificant (Table 2). (*c*) In Experiment II, the mean differences in interference between Groups C1 and C2, and between Groups C2 and C3, representing, respectively, zero and 10, 10 and 20, days' delay between O.L. and I.L., were insignificant whereas the difference between Groups C1 and C3 was significant at the one per cent level. Hence the chances are borderline that 15 days without practice between O.L. and I.L. would significantly affect interference (Table 4). (*d*) In Experiment III, none of the intergroup mean differences in interference, taking the experimental groups in pairs, was statistically significant. Here, up to 15 days' delay prior to R.L. did not appreciably affect interference (Table 4).

TABLE 1
PER CENT RETENTION AND PER CENT INTERFERENCE
($N=10$)*

Experiment	Group	Per Cent Retention			Per Cent Interference		
		First-choice errors	Trials	Times (secs.)	First-choice errors	Trials	Times (secs.)
I	B	82.2	50.7	67.7	−2.4†	−14.7†	−9.9†
	C	58.5	31.6	52.6	30.9	34.9	22.1
	D	71.8	36.4	63.5	15.2	24.9	5.9
	E	71.0	34.7	61.9	16.2	28.5	8.3
	Con. 1	84.7	48.5	67.5			
	Con. 2	80.2	44.2	61.6			
II	C1	55.9	37.9	51.0	26.4	23.6	17.7
	C2	59.8	35.0	56.7	21.2	29.4	8.7
	C3	67.1	44.4	56.4	11.6	10.5	9.2
	Con.	75.9	49.6	62.1			
III	C1	68.4	46.4	70.2	22.6	−5.5†	3.4
	C2	63.8	37.0	68.3	27.8	15.9	6.1
	C3	64.0	38.5	68.9	27.6	12.5	5.2
	C4	57.0	31.5	60.7	35.5	28.6	16.6
	Con.	88.4	44.0	72.7			

* Groups B and E, Experiment I, had 9 rats each; Groups C1, C2, and the controls, Experiment III, had 11 each.

† A minus sign indicates facilitation.

Over-all Retention and Interference

The group results were treated by the *savings method* (Table 1). Retrace errors have been omitted from this table because of their great variability and the small amount of interference they revealed. In the first experiment the O.L. task was retained best by the controls, then by the experimental groups in the order, B, D, E, and C. Marked interference occurred for Groups C, D, and E, in comparison with the controls, and this was greatest for Group C on every index. The interpolation of Pattern B facilitated the relearning of Pattern A. The data for Experiments II and III, in which Pattern C was the only interpolated task, also reflect marked interference. In Experiment II, interference tended to decrease with increase in temporal separation of the I.L. and O.L. tasks. In Experiment III, it tended to increase with increase in number of I.L. trials. The maximum group interference in these experiments, in terms of first-choice errors, was 35.5 per cent, the mean for the ten experimental groups, 23.5 per cent. Such results clearly depend on specific experimental conditions and the present ones may not have been optimal. However, they produced a much smaller amount of interference than that typically shown by human subjects in verbal tasks.

The relative positions of the group mean first-choice error curves (Fig. 1) indicate that there was net positive transfer during I.L., and to a lesser extent during R.L., except for Group C. This is most marked for Group B. The control group curves are similar in form and about the same distance apart. The R.L. curves for the controls are closer to the abscissa than the other R.L. curves, except that of Group B. The relative positions of the I.L. and R.L.

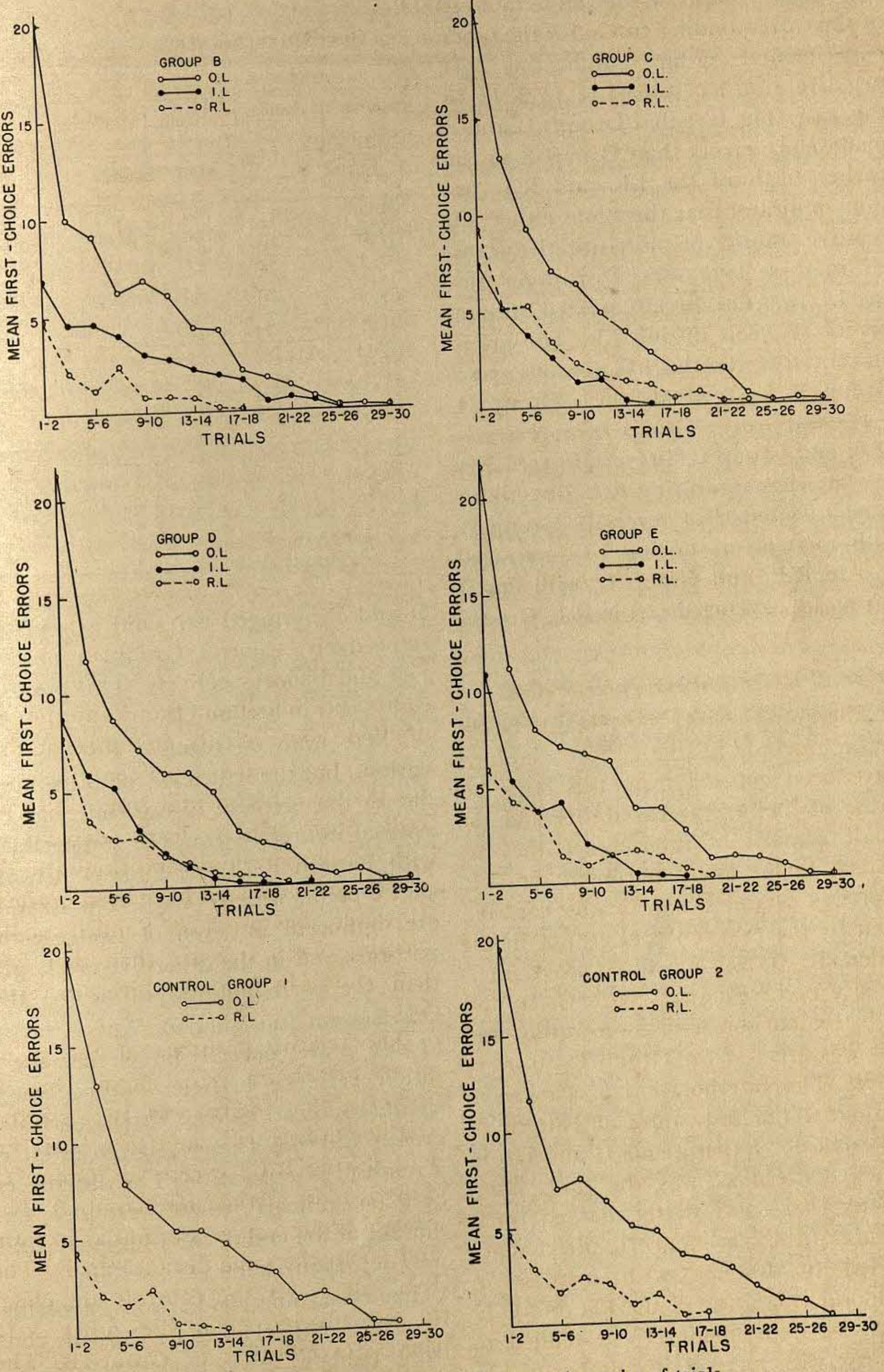


FIG. 1. Mean first-choice errors for successive pairs of trials.

curves for Group C are opposite to those of the corresponding curves for the other experimental groups, indicating net negative transfer from I.L. to R.L. in this case. The fact that Groups D and E made more errors than Group B in the earlier trials of the I.L. and R.L. sessions indicates that there was either less positive transfer or more negative transfer in these two cases. Disruption during I.L. was not directly related to number of reversals during I.L. The group mean errors for the I.L. sessions were: Group B, 14 reversals, 36.2; Group D, 8 reversals, 26.4; Group E, 4 reversals, 28.8; and Group C, 8 reversals, 22.2. Nor was interference during R.L. directly related to number of reversals. Group B, with most errors during I.L., averaged 13.0 in R.L. and Group C, with fewest during I.L., averaged 31.1 in R.L. Groups D and E averaged 21.1 and 21.9 errors respectively, Control Groups 1 and 2, 11.8 and 16.6, respectively. This point is elaborated in Section III.

TABLE 2

SIGNIFICANCE OF GROUP DIFFERENCES IN TERMS OF *t* RATIOS: EXPERIMENT I

Groups	First-choice errors	Retrace errors	Trials	Times (secs.)
	I.L. Session			
B-C	3.27**	2.31*	4.68**	3.18*
B-D	2.28*	2.08	3.99**	1.45
B-E	2.01	1.48	3.74**	1.59
C-D	1.59	.03	.65	1.14
C-E	1.91	1.08	1.47	.96
D-E	.72	.94	.70	.21
	R.L. Session			
B-C	3.77**	.63	1.74	2.29*
B-D	2.64*	.64	1.71	.67
B-E	2.37*	.09	2.19*	.84
B-Control 2	.91	.11	.73	.71
C-D	2.17*	.82	.34	1.97
C-E	1.77	.60	.09	1.56
C-Control 1	3.92**	.61	2.01	2.20
D-E	.22	.73	.53	.41
D-Control 1	3.05**	.59	2.07	.56
E-Control 1	2.69*	.04	2.41*	.88
Controls 1 and 2	1.24	.08	.96	.63

* Significant at the 5% level.
** Significant at the 1% level.

TABLE 3

SUMMARY OF *F* RATIOS: EXPERIMENT I

Source	First-choice errors	Retrace errors	Trials	Times
	O.L. Session			
B.G.	.12	.06	.14	.03
W.G.	1.20	1.34	1.00	1.49
B.R.	2.18*	3.05**	.99	3.96**
	I.L. Session			
B.G.	5.17**	2.56**	12.17**	5.87**
W.G.	1.73	.85	1.16	1.05
B.R.	1.99	.40	1.63	1.22
	R.L. Session			
B.G.	7.10**	.32	2.46*	1.55
W.G.	1.13	.94	1.19	1.11
B.R.	1.79	.66	2.13*	2.16

* Significant at the 5% level.
** Significant at the 1% level.

There were no controls for the I.L. sessions, but the nature of the disruption due to the presence of reversals may be assessed indirectly from the comparisons with Pattern B (Table 2) in which all choice points were reversed. The *t* ratios are significant in seven of twelve comparisons, and in the other five are larger than the corresponding ratios for the O.L. session (not shown). The *F* ratios (Table 3) are generally larger than the *t* ratios, variance between groups being significant on all indices. Variance within groups did not change significantly as a result of interpolation. The differences between animals (rows) were not significant at the end of the I.L. session and were less than in the O.L. session.

The relearning comparisons involving Group B, other than those with the controls (Table 2), are not relevant to interference, but they indicate the change from I.L. to R.L. in the distribution and magnitudes of the significant *t* ratios among the performance measures. In the R.L. session, trials did not generally differentiate the groups in terms of *t* ratios and the *F* ratio for this index is much smaller. Retrace errors did not differentiate the groups. This kind of error decreased markedly during relearning; within the groups its mean varied greatly from trial to trial and the total group mean differences were small. First-choice errors provided the only useful index of interference for experimental-control group comparisons. The variance between groups on this index increased from I.L. to the end of R.L. For these reasons, only first-choice errors were employed in the subsequent analyses of the effects in the separate units.

In Experiment II, interference was significant in two of the comparisons with the controls but in the case of only one of the other group differences, C1-C3 (Table 4). Interference then, decreased as the temporal interval between O.L. and I.L. increased, but it required from 10 to 20 days of practice to differentiate the experimental groups significantly. The *F* ratios were not computed for these data.

TABLE 4

SIGNIFICANCE OF GROUP DIFFERENCES IN FIRST-CHOICE ERRORS IN THE R.L. SESSIONS: EXPERIMENTS II AND III

Groups	*t* ratios	
	Exp. II	Exp. III
C1-C2	1.00	.10
C1-C3	2.95**	.76
C1-C4	—	.96
Cl-Control	3.92**	5.81**
C2-C3	1.04	.82
C2-C4	—	1.00
C2-Control	2.94**	7.19**
C3-C4	—	.39
C3-Control	1.87	7.00**
C4-Control	—	5.57**

** Significant at the 1% level.

All four groups in Experiment III made significantly more errors during R.L. than the controls (Table 4). While interference tended to increase to a maximum with increase in amount of interpolated practice, then decreased with further practice, none of the mean differences between the experimental groups was statistically significant.

Analysis of the Response Pattern

The O.L. curves (Fig. 2) show the typical pattern of unit errors made by these groups. The curves reflect a tendency to backward elimination of errors partly obscured by a marked tendency to enter the last goal-pointing blind in Unit 12 and an uptilt in the curve toward the goal due to relatively more entrances to the blinds in Units 10, 11, 13, and 14. These trends are reflected in a convex serial position curve of errors which is the reverse of that commonly found in serial verbal learning, probably due to the fact that errors of anticipation and perseveration in this maze could produce correct responses, especially in the units near the center. In general, errors were relatively more numerous in the even-numbered units and relatively fewer in the odd-numbered ones, in Units 4 to 14, inclusive. The fact, however, that this trend was disturbed between Units 2 and 4 (3 and 5 in the case of Pattern D), suggests that there were opposing response tendencies in this area. Incidentally, this feature does not appear in the corresponding curves for retrace errors (not shown); all of them exhibit the regular alternating pattern throughout.

A further analysis of the O.L. response pattern showed that: (*a*) there was no preference for either side of the maze,

Bureau Ednl. Psy. Research
ING COLLEGE
Dated

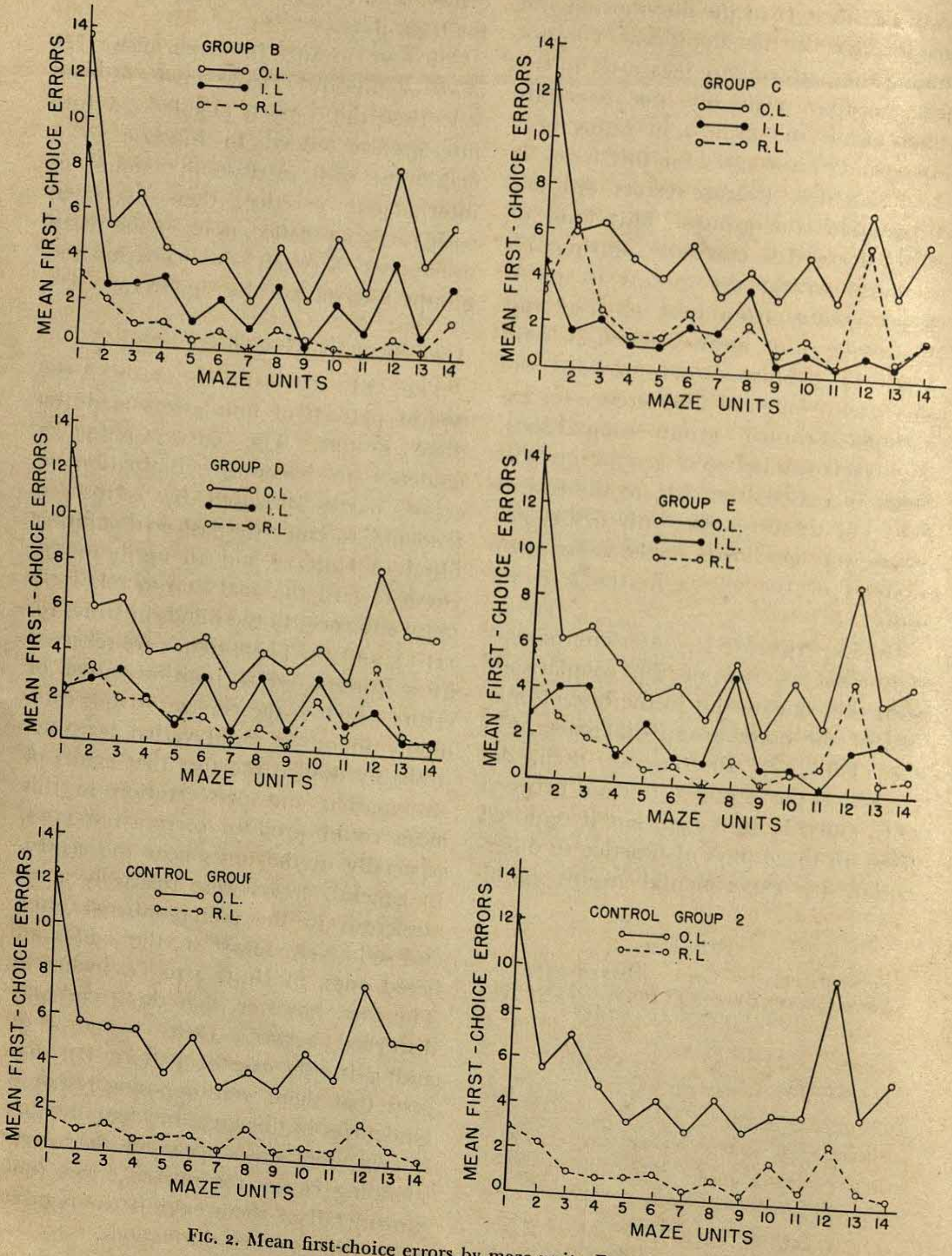


FIG. 2. Mean first-choice errors by maze units, Experiment I.

disregarding errors in Units 1 and 12; (*b*) disregarding the errors in Units 1, 2, 12, and 14, fewest entrances were made into blinds pointing backward, i.e., away from the food box at a 90-degree angle with it; (*c*) disregarding Unit 12, the blinds in Units 6, 10, and 14, pointing directly away from the food box, were entered about as frequently as the other even-numbered blinds; (*d*) there was no appreciable indication of handedness in the data. The most marked response tendency was the alternating pattern of errors jointly constituted by the responses described in (*b*) and (*c*).

The effects of interpolation on the original pattern of errors could not have been predicted from the number and serial positions of the reversals. These changes did not result in a generally different response pattern, but any subsequent disruption of behavior radical enough to reverse the direction of the error curve tended to occur near the start of the maze and to a lesser extent near the goal. The pattern of errors in the other units was quite stable, even when the absolute numbers of errors in specific units varied greatly from one pattern to another and from O.L. to I.L. for a given pattern.

In more detail, the I.L. curve for Pattern B, with all units reversed, is generally similar in form to the O.L. curve for Pattern A, except that the curve changes direction between Units 2 and 3, instead of 3 and 4. The I.L. curve for Pattern C, with reversals in Units 2, 3, 6, 7, 8, 9, 12, and 13, differs only slightly from that for Pattern A, namely, in the number of errors in Unit 4 relative to Unit 5. The reversals in Pattern D were at the ends of the maze in Units 2 to 5, inclusive, and in Units 10 to 13, inclusive, but the disturbance of pattern was near the start. The I.L. curve for Pattern E, with reversals in Units 1, 3, 12, and 14, changes direction relative to the O.L. curve between Units 1 and 2, 5 and 6, and 13 and 14. Here, a reversal in Unit 1 did not produce maximum errors in Unit 1 but in Unit 2, whereas in the curve for Pattern D the same thing occurred although Unit 1 was not reversed during I.L. In the Pattern B curve, a reversal in Unit I did not produce a change in the frequency of errors in this Unit. There were no reversals in Units 6 to 9, inclusive, of Patterns D and E, but the frequencies of errors differ markedly in these Units in the two patterns, although the general forms of the curves are similar.

In the absence of controls for the I.L. session, the statistical significance of the above changes can be assessed only indirectly by reference to Pattern B. As one of a number of analyses, the *t* ratios were computed for the differences in first-choice errors in each maze unit between the experimental groups taken by pairs, the purpose being to provide a comparison of the pattern of significant disruption during I.L. with that of interference during R.L. None of the corresponding differences between the O.L. group scores in each unit, taking the groups two at a time, was significant. In the I.L. data, there were significant differences in 16 of 42 unit-pair comparisons involving Pattern B. But significant differences failed to appear in 11 of 22 comparisons with Pattern B in which only one of the choice points in the patterns being compared was reversed during I.L., and they did appear in scores in 5 of 10 units in which both choice points in the two patterns being compared had been reversed. Together with the foregoing description of the O.L. and I.L. curves, these data indicate in a general way the extent of disruption due to interpolated reversals and the limits of specificity of the effects of reversals on disruption.

The R.L. curves should be read with the *t* ratios for the mean group differences between the experimental and control group scores in each maze unit (Table 5). The positions of the interpolated reversals are indicated by lines under the *t* ratios for the units affected. Interference occurred mainly at or near the start of the maze in Units 1 and 2, and also in Unit 12. These units accounted for seven of twelve significant differences, Units 4, 6, and 14 for the other three. Despite the diverse patterns of interpolated reversals, the forms of the R.L. curves are very similar to that for Pattern A. Greater dispersion of the errors indicative of interference would have been expected if the effects of reversals had been specific to their loci during I.L. The only marked departures in the experimental group curves from the corresponding points in the O.L. curves are (*a*) a shift of the point or points of reversal of direction of the alternating error curve (B,C, D, and E), and (*b*) displacement of the unit of maximal error frequency near the start of the response pattern from Unit 1 to Unit 2 (C and D). However, changes similar to (*a*) occurred in the curves for the control groups, which had had no I.L.

As for positions of reversals, in Pattern B, with all choice points reversed during I.L., the rats continued to make errors maximally in Unit 1 during R.L. In this case, however, all the unit-score differences are insignificant except that for Unit 12, which is significant for facilitation. In Pattern C, Unit 1 was not reversed during I.L. but Unit 2 was; here, however, relatively more errors were made in Unit 2 than in Unit 1 during R.L., and significant interference also

TABLE 5

SIGNIFICANCE OF DIFFERENCES IN FIRST-CHOICE ERRORS BY MAZE UNITS BETWEEN EXPERIMENTAL AND CONTROL GROUPS: EXPERIMENT I, R.L. SESSION

Group Pair	Maze Unit													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
B-Control 2	**.08**	**.26**	**.00**	**.29**	**.54**	**.19**	**1.82**	**.11**	**.33**	**1.57**	**.91**	**4.32****	**.34**	**1.75**
C-Control 1	2.37*	**4.70****	**.32**	1.04	.95	**3.20****	**.28**	**1.18**	**.47**	1.15	.37	**4.38****	**.00**	2.26*
D-Control 1	1.19	**2.76***	**1.06**	**2.21***	**1.27**	.68	.00	.46	.26	**1.63**	**.13**	**1.72**	**.20**	.26
E-Control 1	**3.65****	1.90	**1.05**	1.23	.29	.21	2.00	.26	.29	.10	.52	**3.60****	.27	**.65**
Controls 1 and 2	1.24	1.16	.35	.34	.18	.17	.84	.73	.14	1.17	.37	1.20	.00	.15

Note.—Boldface indicates a reversal during I.L.
* Significant at the 5% level.
** Significant at the 1% level.

occurred in Units 6, 12, and 14, all reversed. In Pattern D, there was a similar effect to that for Pattern C in Units 1 and 2, but with relatively fewer errors in Unit 2, and a significant amount of interference also occurred in Unit 4 (reversed). In Pattern E, with a reversal in Unit 1 during I.L., there was an effect during R.L. similar to that in the case of Pattern B. In the three patterns that produced interference, therefore, when interpolated reversals came in Unit 2, they were accompanied by maximal interference in Unit 2 rather than Unit 1. Reversals in Unit 12 were also followed by interference (except for Pattern B) which was significant in the cases of Patterns C and E, but not by a change in the direction of the curve. This may have been due, of course, to the fact that Unit 12 was a major source of errors in all sessions.

Except in the case of Unit 6, Pattern C, and Unit 4, Pattern D, interpolated reversals in units other than the first two and No. 12 were not accompanied by a significant number of errors indicative of interference nor changes in the direction of the curve during R.L.

The patterns of *t* ratios for the differences between the unit scores for I.L. and R.L. were also compared, taking the four experimental groups two at a time. This analysis is obviously of no relevance to absolute interference since the controls were not involved. However, it brought out clearly the lack of correspondence between the pattern of response disruption during I.L. and that indicative of interference in R.L. There were 30 significant *t* ratios in the I.L. comparisons, twice as many as in the R.L. comparisons, but only for three units were they common to both sessions, namely, two group-pair comparisons for Unit 2, one for Unit 10, and one for Unit 12.

Interference tended to occur at loci of reversals rather than loci of nonreversals—there were seven and two significant *t* ratios, respectively—but it failed to occur significantly at 13 loci of reversals, excluding the 14 in Pattern B. Reversals tended to produce relatively more interference when they were the first of a series. This occurred in Pattern C, Units 2, 6, 12; in Pattern D, Unit 2; and in Pattern E, Units 1 and 12, in the last case a series of two reversals interrupted by a nonreversal. Spread of the interference effect of reversals is also indicated by the fact that nonreversals immediately adjacent to reversals tended to result in greater score differences between the experimental animals and the controls than at other choice points, although not generally in terms of statistical significance. This happened in Pattern C, Units 1, 4, 5, 10, and 14; in Pattern D, Units 1 and 6; and in Pattern E, Units 4 and 11. The exceptions to this trend are: Pattern C, Unit 11; Pattern D, Units 9 and 14; and Pattern E, Unit 13.

The total error curve by maze units for Trials 1 and 2, and the means for the first two pairs of trials, the first three pairs, etc., to and including Trials 1 through 12 were plotted for the O.L. session; the total errors for Trials 1 and 2 and the means of the first six pairs of trials only were plotted for the other two sessions because of excessive fluctuations in the curves for these sessions. These curves are not shown here. This analysis revealed that the response pattern was clearly identifiable in the early series of trials and changed little as practice proceeded. The curve tended to be flat for the initial pair of trials and the bowing became progressively more marked with practice.

There are two further points of note.

First, the total errors in Unit 1 was 108 for the first two trials; the average of the first two pairs was 106, and of the first three pairs, 134. After this the means decreased consistently. This did not happen for any other unit; in all of them the means decreased from the start as the total errors for successive pairs of trials were included in the average. The persistence of errors in Unit 1 relative to the errors in units nearer the goal would be expected in terms of a goal gradient effect, but hardly an increase in the cumulative average for the first six trials in a maze being encountered for the first time. Second, there was a tendency for the regular alternating pattern of the curve to shift backward toward the start of the maze from the earlier to the later trials. The curve for Trials 1 and 2 changes at Unit 6, and that for Trials 1 through 12 at Unit 4. Since the O.L. pattern of errors also alternates between Units 1 and 2 and Units 2 and 3, this may be looked at another way: as a tendency for the change in the expected direction of the alternating curve to appear closer to the start of the maze in the later stages of practice. The pictures for the I.L. and R.L. sessions are not as clear as for the O.L. session because of previous practice. The curves were distinctly bowed from the start of running, and the final characteristic alternating pattern of errors was in evidence at all stages of practice.

III. DISCUSSION

It may be stated, to a large extent retrospectively, that the heuristic use of the similarity factor in the first experiment could not be expected to enlarge our knowledge of the relationship between interference and similarity as defined by the patterns used here because, in addition to number and serial positions of interpolated reversals, there were factors associated with task complexity, such as changes in the line of progress and asymmetrical arrangement of the entrance and goal boxes, which were not defined operationally at the outset nor varied systematically. This point becomes quite clear when the inverse relation between interference and difficulty of the I.L. task is submitted to a more elaborate test of the applicability of the similarity factor than was available to Webb (19), who obtained a similar result in 1917. Let it be assumed that the rat, on approaching choice points, always met the same field of sensory stimulation in O.L and I.L., but that, during I.L., certain choice points required different responses than during O.L. The assumption is, of course, questionable for motor activity since it takes no account of the role of kinesthesis associated with sequence of turns. Osgood's (14) three subparadigms, based on his synthesis of the verbal-learning research on the similarity factor, predict the kind and direction of transfer and interference to be expected, presumably in all cases of variation of stimuli and responses. Only the second subparadigm appears to apply here, which predicts, for the case in which stimuli are functionally identical in the second task of the transfer series and responses vary along a dimension of similarity, that there will be negative transfer and interference, both decreasing with increased similarity between the responses. It is clear from Table 6 that the relation between interference and difficulty of the interpolated task is not a simple function of the number of reversals. If, following Ballachey's results (1), Pattern B were regarded as functionally most similar to Pattern A, then interference would vary inversely with degree of similarity between O.L. and I.L. in accordance with Osgood's subparadigm. But there still remains the task of explaining why Pattern B became functionally most similar to Pattern A in terms that permit prediction since it was the most difficult interpolated task, and also why the opposite was true for Pattern C. The error curves (Fig. 1) show that Pattern B did not produce the greatest mean error until after the fifth or sixth trial, prior to which the group means were about equal. The greater difficulty of Pattern B during I.L. was not uni-

TABLE 6

COMPARISON OF DEGREE OF INTERFERENCE, NUMBER OF REVERSALS, AND DIFFICULTY OF THE TASKS

Pattern	B	C	D	E
t ratios of interference	.41	3.92	3.05	2.69
No. of reversals	14	8	8	4
Mean error in I.L.	36.2	22.2	26.4	28.8
Mean error in R.L.	16.6	31.1	21.1	21.9

formly distributed over the reversals, but presumably was due largely to some special function associated with Unit 1, for 67 per cent of the possible errors in the first eight trials occurred here. The corresponding percentages for Patterns C, D, and E are 56, 28, and 26. To a lesser extent, the same thing occurred in Unit 12, the percentages being 26, 10, 21, and 22 for Patterns B, C, D, and E, respectively; and in Unit 14, the percentages being 29, 18, 9, and 17, respectively, for these patterns. The same thing cannot be said for any other unit in Pattern B. Similar analyses of the R.L. data showed that the blinds in Units 1, 2, and 12 were much more potent sources of interference than the other blinds, and that these units may be regarded as the focal points of patterned responses. The foregoing merely demonstrates the limitations of the transfer theory of interference when stimuli and responses are defined as they were in these complex tasks. The difficulty seems to be primarily one of keeping the stimuli and responses operationally discrete in motor learning.

The original response pattern may be explained by the following principles: (*a*) a goal gradient effect (10) for the tendency to backward elimination of errors; (*b*) either *centrifugal swing* (1, 15) or *forward-going tendency* (3) for the alternating pattern of errors reflected in relatively more errors in the even-numbered units, whether or not they were goal-pointing; both these principles seem to be reflected by the data and both express the effects of local maze structure; (*c*) goal anticipation (2, 16, 17) or *choice-point expectancy* (6) for the predominance of errors in Unit 12 and to a lesser extent in the immediately preceding units and the general uptilt in the part of the curve near the goal, excluding Unit 14. (*d*) The operation of a hypothetical principle of abstractive behavior (12), involving set formation or positioning, would seem to be necessary to account for the behavior at the start of the maze, especially the predominance of entrances to the blind in Unit 1. It would seem, moreover, that such a principle encompassed not only the general direction of the food box, but also the alternating pattern of responses. It is not essential that these principles be characterized more exactly. The data clearly seem to support the generalization that interference, rather than being tied in with the specific loci of interpolated reversals, tended to occur in units in which, during original learning, errors tended to occur more frequently where they would be expected to occur in accordance with these principles as far as they can be identified.

A few further details of the pattern analysis may be cited in justification of the above conclusion. The single alternation Pattern C was easiest during I.L., but produced most interference with relearning. Since it contained eight reversals in comparison with 14 in Pattern B and eight in Pattern D, an explanation of the result was then sought in terms of the principles of centrifugal swing or forward-going tendency. This showed that the pattern of interference, including increases in errors which were not statistically significant, increased and decreased from unit to unit in almost parallel fashion with the pattern of first-choice errors in O.L., regardless of the loci of reversals in I.L. (Table 7), and both patterns of errors can be accounted for largely in terms of these principles. The parallelism does not hold for Units 2 and 3, the area in which the curve of errors changed direction. This circumstance then seemed to call for the operation of a principle of responding that was different from any yet considered. A similar parallelism can be demonstrated for Pattern D, but it is difficult to do so for Pattern E in view of the more complex arrangement of turns. Pattern B, of course, produced facilitation, but the fact that the negative transfer produced during I.L. by this pattern was concentrated mainly in Units 1, 12, and 14, and also its functional similarity to Pattern A during R.L., would also appear to be evidence in favor of some sort of abstractive behavior.

The predominance of errors in Unit 1 during O.L. is understandable in terms of backward elimination of errors, but this principle does not account for the increase in errors in

TABLE 7

COMPARISON OF PATTERN OF O.L. ERRORS AND PATTERN OF INTERFERENCE: GROUP C

Unit	O.L.		R.L.	
	Turn	Mean Error	Turn	*t* ratio (Interference)
1	L	12.6	L	2.37*
2	L	5.9	**R**	4.70**
3	R	6.3	**L**	.32
4	R	4.8	R	1.04
5	L	4.0	L	.95
6	L	5.5	**R**	3.20**
7	R	3.3	**L**	.28
8	R	4.4	**L**	1.18
9	L	3.2	**R**	.47
10	L	5.1	L	1.15
11	R	3.2	R	.37
12	R	7.1	**L**	4.38**
13	L	3.6	**R**	.00
14	L	5.9	L	2.26*

Note.—Boldface indicates a reversal during I.L.

* Significant at the 5% level.

** Significant at the 1% level.

this unit over the early trials nor for the intriguing shift of the unit of maximal errors to Units 2 and 3 (Patterns D and E) during I.L. which, as already shown, cannot be attributed to the presence of reversals. The blinds in the first three units may have served to activate an anticipation or set for both the goal direction and the alternating running response, presumably a set for the correct pattern of turns, the characteristic pattern of errors being forced by local structural features of the maze. In order for centrifugal swing to produce errors, the maze yet be learned, it seems necessary to assume that the behavior at and near the start of the maze was governed by a different principle, one that was relatively independent of situational factors. In the case of Pattern A, the last two correct turns into the food box are left turns so that, in accordance with the principle of reinforcement, turning left in Unit 1 should have become easier in subsequent trials; whereas there was a greater tendency to turn right in the early trials, and this increased for at least six trials. It is as though a bidirectional gradient of reinforcement from one trial to the next began to operate early in O.L. but was offset by some other tendency that caused Blind 1 to be entered more frequently until some other adjustment had taken place. Association by contiguity does not help here if by "the last occasion" is meant either one or two left turns into the food box.

Concerning the reversal of direction of the curve of errors in Units 2, 3, 4, or 5, if the alternating pattern of correct responses had been continued regularly from trial to trial, opposed response tendencies would have been produced on alternate trials; i.e., in Trial 2 the correct turns in Units 1 and 2 in Pattern A would have been RR. It would seem to be necessary, therefore, to assume that, while the pattern of running was forced by the maze structure, it was also learned as a pattern, or "cognitively." The rat, having made two left turns into the food box on a previous run, tended on the next run to make two right turns at choice points which led it into Blinds 1 and 2 in accordance with the alternation pattern of response, but the principle of centrifugal swing which was mainly responsible for the alternating pattern of errors would have produced maximum errors in the even-numbereed blinds, beginning with Blind 2. Therefore, in order for Blind 1 to persist as a potent source of errors in accordance with the assumption that it was the focal point of operation of a different principle while the rat learned the maze, the alternating pattern of errors had to be reversed somewhere along its course. This tended to occur near the start of the maze, with the suggestion that it occurred nearer the goal in the early trials. This point needs to be tested further, however, possibly employing massed practice and finer units of measurement.

The argument, so far, seems to justify the assumption of a principle of responding in addition to backward elimination of errors to account for the errors in Unit 1 during O.L. A consideration of the state of affairs during I.L. and R.L., however, does not permit a more precise specification of the nature of such a principle other than that the responses which it would have to encompass would require that it be one of considerable lability. During I.L., Units 2 and 3 in Patterns D and E displaced Unit 1 as the maximal sources of errors, and there were similar shifts during R.L. which are not ascribable to the presence of reversals during I.L. During I.L., the occurrence of a first reversal in Unit 2 in Patterns C and D was attended by opposite effects—in the first case no change in the status of this unit as the one in which most errors were made, and in the second case the emergence of Unit 3 as the greatest source of errors. During R.L., in the case of Pattern C, Unit 1 lost its status to Unit 2 as the maximum source of errors, but in the case of Pattern D the effect that occurred in I.L. persisted through R.L.

This analysis is too descriptive to establish beyond doubt the case for a principle of abstractive behavior. Questions must be answered as to why the rat must actually enter the blind in Unit 1, rather than merely pausing, in order to get a hypothetical set functioning, and why, indeed, it did not become functional in the entrance box or first alley—but they are beyond the scope of this study.

The results of the first experiment weaken the case for the transfer theory of interference based on verbal-learning research to the extent that the transfer was not a graded function of similarity, and to the extent that determinants of interference associated with pattern nullified the influence of specific reversals. It may, of course, be possible to express patterned effects in terms of similarity. The finding in the second experiment of most interference for interpolation adjacent to O.L. is contrary to most of the findings in verbal-learning experiments which have indicated that there are other high points for interference nearer to R.L. or that interference is independent of the temporal position of the interpolated task. The present outcome could therefore be taken as supporting the somewhat outmoded perseveration theory of interference if a neurophysiological setting-in process operating in the order of days, rather than seconds, could reasonably be assumed to underlie the retention of the O.L. task. The results of Experiment III are inconclusive.

Much remains to be done to define the relationships between the principles of maze running and interference. The limiting conditions of the effects on interference of factors associated with response pattern, in comparison with the effects of reversals, need to be examined with mazes of fewer units and of a more comprehensive experimental design. All possible variations between the O.L. and I.L. tasks in congruent and adjacent units should be imposed. For example, it would be interesting to test for spread of interference at loci of reversals that were followed and preceded by true paths leading in opposite directions. Interference for mazes in which the different response patterns were produced in uncomplicated forms might be compared. In making such comparisons it would be of interest to specify the genesis of facilitation and interference in such terms as the emergence of dominant units in the production of these effects. The problem of the extent to which interference is a "cognitive" or an emotional function might be explored in terms of comparisons of the results for first-choice errors with those for other recorded performance indices. In the present study, the divergent results during I.L. and R.L. for the different indices of performance may mean that transfer and interference are mediated by different processes.

IV. Summary

A fourteen-unit, multiple-T maze was the original and relearned task for matched groups of 10 rats in three experiments in which the interpolated tasks were, respectively, (*a*) four patterns differing in number and loci of reversed choice points, (*b*) the task producing most interference in the first experiment, practiced at three temporal points, and (*c*) 5, 10, 15, or 20 trials in the same task as in *b*. Significant interference was shown by 9 of the 11 experimental groups, most markedly in the first two R.L. trials, but until at least the eighth trial. The tenth group showed insignificant interference; the eleventh, significant facilitation. The mean group interference was 23.5 per cent; the maximum, 35.5 per cent. Interference varied inversely as the difficulty of the I.L. task. It decreased with increase in temporal separation of the I.L. and O.L. tasks, and increased as the mean number of trials increased to a point about equal to the mean O.L. trials, after which it decreased. In the latter case, the changes in amount of interference were not statistically significant. The typical serial learning curve, reflecting the operation of several known principles of maze running, was established early in the O.L. series and persisted through the other series despite the interpolation of diverse patterns of reversals. Interference was not specific to the loci of reversals, but tended to occur at points in the maze where the errors in O.L. were attributable to the identified principles, particularly near the start and in the last goal-pointing blind.

REFERENCES

1. Ballachey, E. L., & Buel, J. Centrifugal swing as a determinant of choice-point behavior in the maze running of the white rat. *J. comp. Psychol.*, 1934, **17**, 201-223.
2. Ballachey, E. L., & Buel, J. Food orientation as a factor determining the distribution of errors in the maze running of the rat. *J. genet. Psychol.*, 1934, **45**, 358-370.
3. Bayroff, A. G. Direction orientation and forward-going tendency in white rats. *J. comp. Psychol.*, 1933, **15**, 211-218.
4. Bogardus, E. S., & Henke, F. G. Experiments on tactual sensation in the white rat. *J. anim. Behav.*, 1911, **1**, 125-137.
5. Britt, S. H. Retroactive inhibition: A review of the literature. *Psychol. Bull.*, 1935, **32**, 381-440.
6. Buel, J. The linear maze. I. "Choice-point expectancy," "correctness," and the goal-gradient. *J. comp. Psychol.*, 1934, **17**, 185-199.
7. Cooper, J. B. The effect upon performance of introduction and removal of a delay within the maze. *J. comp. Psychol.*, 1938, **25**, 457-462.
8. Gengerelli, J. A. The principle of maxima and minima in animal learning. *J. comp. Psychol.*, 1930, **11**, 193-236.
9. Gilhousen, H. C. Temporal relations in anticipatory reactions of the white rat in a maze. *J. comp. Psychol.*, 1938, **26**, 163-175.
10. Hull, C. L. The goal gradient hypothesis and maze learning. *Psychol. Rev.*, 1932, **39**, 25-43.
11. Kogan, L. S. Interference in maze learning as a factorial function of similarity and goal gradient. *J. exp. Psychol.*, 1951, **42**, 69-81.
12. Krechevsky, I. "Hypotheses" in rats. *Psychol. Rev.*, 1932, **39**, 516-532.
13. MacCorquodale, K. The effects of pattern interruption on maze performance. *J. comp. Psychol.*, 1942, **33**, 363-369.
14. Osgood, C. E. Meaningful similarity and interference in learning. *J. exp. Psychol.*, 1946, **36**, 277-301.
15. Schneirla, T. C. Learning and orientation in ants. *Comp. Psychol Monogr.*, 1929, **6**, No. 4.
16. Spence, K. W. The order of eliminating blinds in maze learning by the white rat. *J. comp. Psychol.*, 1932, **14**, 9-27.
17. Spragg, S. D. S. Anticipation as a factor in maze errors. *J. comp. Psychol.*, 1933, **15**, 313-329.
18. Spragg, S. D. S. Anticipatory responses in the maze. *J. comp. Psychol.*, 1934, **18**, 51-73.
19. Webb, L. W. Transfer of training and retroaction. *Psychol Monogr.*, 1917, **24**, No. 3 (Whole No. 104).
20. Wiltbank, R. T. Transfer of training in white rats upon various series of mazes. *Behav. Monogr.*, 1919, **4**, No. 1.

(Accepted for publication October 17, 1955)

Vol. 70, No. 9 Whole No. 416, 1956

Psychological Monographs: General and Applied

Studies of Independence and Conformity:
I. A Minority of One Against a Unanimous Majority[1]

Solomon E. Asch
Swarthmore College

I. INTRODUCTION

THE investigations described in this series are concerned with the conditions of independence and lack of independence in the face of group pressure.[2]

Of the many diverse forms of social independence and submission, we have selected one in particular for study. By means of a procedure shortly to be described we generated a disagreement between a single person and a group concerning a simple and clear matter of *fact* in the immediate environment. Further, the group that disagreed with the individual judged the facts in question *wrongly*, while the individual could not but judge

---

[1] This is the first of a series of reports describing an extensive investigation of group pressures and their effects on judgment. The scope of the studies required support and the help of many persons.

I am glad to record my gratitude for financial assistance from the Office of Naval Research, which supported these investigations as part of its policy of encouraging basic research in the psychological disciplines. In particular I wish to extend my thanks and appreciation to Dr. J. V. Macmillan and Dr. Howard E. Page of the Office of Naval Research for their helpfulness and for their devotion to the interests of science.

In the conduct of the investigations I was particularly fortunate to have the assistance of a number of psychologists whose cooperation and enthusiasm made the work possible. The reader of these pages will soon discover that we were engaged in exacting experiments which can be justified on the ground of their potential scientific value provided the investigator treats the subject with respect, and succeeds in conveying to him that he is making a contribution. In such matters one can only rely on the sensitiveness and human feeling of the experimenter. I am proud to say that those who were associated with me in this work fully justified this confidence. It is with pleasure and thankfulness that I mention the help in the conduct of the experiments of Dr. David A. Emery, Miss Esta Soloway, Mrs. Enid Hobart Campbell, Dr. Dorothy Dinnerstein, Dr. Irwin M. Rosenstock, Mr. Jack Hahn, Mrs. Lillian Z. Berg and Mr. Irwin Feinberg. To Dr. Henry Gleitman I am indebted for expert help in the designing of many experiments in this series and for the statistical analysis of the data. In the latter work he was also ably assisted by Mr. Robert H. Peters.

Mrs. Doris M. Joseph provided the arduous secretarial assistance. In time she also became responsible for the innumerable practical problems accompanying an extensive undertaking. These tasks required, in addition to skill, a considerable expenditure of good spirits. For her help in keeping our work on an even keel I am very grateful.

The studies were conducted in three institutions of higher learning. It seems proper not to identify them by name. This decision deprives me of the opportunity to acknowledge specifically the friendly cooperation of the authorities in these institutions, but I hope they will understand that their contribution has not been forgotten.

It remains only to mention the many young persons who served as subjects in these experiments and who must also remain anonymous. If any of them should happen to read these pages they will, I hope, realize what we tried to have them understand at the time, namely, that the work in which they took part was not only with and about them, but also for them.

To prevent a possible misunderstanding in unwary readers I might also mention that among the subjects of the present series of studies there were no Swarthmore College students.

[2] The present study grew out of earlier experiments that have been reported in the writer's *Social Psychology*, Chapter 16 (1). A preliminary abstract of the present studies has appeared in Asch (2).

Bureau Ednl. Psy. Research
DAVID HARE TRAINING COLLEGE
Dated ....

the facts correctly. Finally, the judgments were stated *publicly;* the single individual was always called upon to announce his judgment just after a group of equals had stated a wrong judgment. In short, we are concerned with public independence and lack of independence in the face of arbitrary group opposition. The aim was to observe the impact of these conditions when the question at issue was that of resisting or bowing to a prevailing group direction. More generally, the object of the present investigation and of those to follow is to give an account of the facts observed and to state some of the conditions responsible for independence and failure of independence.

This is therefore a study of a particular problem in the wide region of social influences. While the fact of social influence is beyond doubt, we are only on the threshold of understanding the responsible processes. The task of inquiry in this region is to explore the ways in which group actions become forces in the psychological field of persons, and to describe the forces within persons that cooperate with or resist those induced by the group environment. In the early stages of investigation the solution to these questions appeared obvious. The far-reaching compliance of persons with group demands was referred to a psychological tendency to "uncritical acceptance" of group ideas and evaluations. General observation and controlled studies seemed to support the conclusion that the fundamental social-psychological process was that of conformity. But the notion of conformity is essentially a restatement, in the guise of an explanation, of the observable events and adds little to our understanding of them. A more analytical reward-punishment psychology stood ready at hand to convert the operation of group forces into the familiar terms of current learning theory and to refer action in line with group demands to prospects of reward and fear of punishment. The latter formulation, while it possesses a rough plausibility, again does little more than restate what is known in terms of operations that have not been directly studied. These are substantially the directions that thinking has taken. They have guided the steps of investigation from the choice of problems to the interpretation of findings.[3]

The apparent plausibility of these interpretations should not hide the fact that they are abstract and not based on careful observation. Granting the great power of groups, may we simply conclude that they can induce persons to shift their decisions and convictions in almost any desired direction, that they can prompt us to call true what we yesterday deemed false, that they can make us invest the identical action with the aura of rightness or with the stigma of grotesqueness and malice?

The abstract temper of present-day theory and investigation in this region rests to a considerable degree on a neglect of the cognitive and emotional experiences that are part of the individual's psychological field. This accounts for a certain one-sidedness of emphasis and a failure to note distinctions that are obvious enough to common sense. The prevailing accounts have taken as the prototype of social influence an arbitrary and slavish submission to group pressure, committing themselves hastily to a subjectivistic conception that comes near to equating group effects with the production of error and illusion (1, Chap. 14). But we ought to treat with reserve the widespread assumption that there is a single form of social influence which is a prototype for all others. More likely there are diverse forms; like the terms "digestion" and

---

[3] For a discussion of some interpretations of social influence the reader is referred to Asch (1, Chs. 14–16).

"reproduction," "social influence" doubtless refers to a considerable range of processes. In particular, it is not justifiable to assume in advance that a theory of social influence should be a theory of submission to social pressure. One need not doubt the great power of social forces to realize that conformity is not the sole effect they produce. The striving for independence and resistance to encroachment are as much facts about people as is conformity. It is consequently unduly narrowing to emphasize submission, to the neglect of the not inconsiderable powers persons demonstrate on occasion for acting according to conviction and rising above group passion. The understanding of social influences will require the study of a wide range of conditions and of the interrelated operations of different psychological functions.

II. Plan of the Investigation: Experiment 1

A. *General Description*

A group of seven to nine individuals was gathered in a classroom to take part in what appeared to be a simple experiment in visual discrimination. They were instructed to match the length of a given line—the standard—with one of three other lines. One of the three comparison lines was equal to the standard; the other two lengths differed from the standard (and from each other) by considerable amounts. The entire task consisted of 18 such comparisons. Figure 2 shows the main comparisons. The individuals were instructed to announce their judgments publicly in the order in which they were seated. The comparison lines were numbered 1, 2, and 3 from left to right and permitted the subjects to state their judgments by calling out the appropriate number. Table 1 contains the lengths of the standard and comparison lines.

The following condition was the vital feature of the experimental situation. All but one of the group had met previously with the experimenter and were instructed to respond on certain trials with wrong and unanimous judgments. Into this group we introduced a single individual who was not aware of this prearrangement. This individual heard the majority respond unanimously from time to time with estimates that clearly contradicted his own observation, and that departed from the true value by amounts ranging from 3/4 to 1 3/4 inches. For example, on the first comparison reproduced in Figure 2, the majority matched the standard which was 3 inches with a 3 3/4-inch line, or with a line 4 1/4 inches long. That the differences were clearly distinguishable is shown by the fact that under control conditions, namely, with subjects judging individually, the estimates showed an accuracy of over 99 per cent.

This, then, was the essential structure of the experimental situation. By means of an artificial procedure we introduced a sharp disagreement between one person and an entire group when the task was that of judging a clear perceptual relation. We placed a single individual, whom we will call the critical subject, in the position of a *minority of one* against a *wrong and unanimous* majority. Perhaps for the first time this person found a massed majority contradicting the clear evidence of his senses.

B. *Details of Procedure*

The Instructed Majority

The instructed majority consisted of male college students who had volunteered for the purpose. During a training session the general purport of the experiment and their role in it was explained. The majority was instructed to announce the judgments clearly and firmly, but not to take issue with the critical subject. They were also advised not to look directly at him and to refrain from feigning surprise at his answers.[4] The majority was therefore far from militant or aggressive; rather it tended to the side of impersonality.

It might be added that the situation did not call for histrionic talents or any elaborateness of action. One brief explanatory discussion followed by a single rehearsal with one person acting as the minority subject sufficed to initiate the group to the task. It was found advisable, however, to have occasional discussions with the group subsequently for the purpose of correcting certain errors, such as responding too quickly or in too low a voice.

A word need be said about the critical subjects and their introduction to the experiment. They, too, were male college students, drawn from the same population as the majority. A critical subject was always recruited by a member of the majority, the explanation being that a psychological experiment in visual discrimination was about to start for which an additional person was needed. When he arrived he found the group in the room or in the corridor, obviously waiting for the experiment to start. Upon the appearance of the experimenter they took their seats. The critical subject was nearly always seated before the last member of the majority. Unknown to him the majority left the designated seat vacant. This procedure ensured that he would not occupy a conspicuous position and that he would hear the estimates of all but one before it was his turn to speak. (When occasionally the critical subject took another seat, it was not too difficult to suggest a rearrangement that followed the predetermined plan.)

The experimenter opened the proceedings by placing in front of the room the first set of cards and then reading the following instructions:

> This is a task involving the discrimination of lengths of lines. Before you is a pair of cards. On the left is a card with one line; the card at the right has three lines differing in length; they are numbered 1, 2, and 3, in order. One of the three lines at the right is equal to the standard line at the left—you will decide in each case which is the equal line. You will state your judgment in terms of the number of the line. There will be 18 such comparisons in all.
>
> As the number of comparisons is few and the group small, I will call upon each of you in turn to announce your judgments, which I shall record here on a prepared form. Please be as accurate as possible. Suppose you give me your estimates in order, starting at the right in the first row, proceeding to the left, and then going to the second row.[5]

By reading the instructions to the assembled group the experimenter conveyed the impression that all were equally new to the situation. To strengthen this impression some members of the majority asked questions intended to "clarify" the instructions. They inquired whether there would always be a comparison line equal to the standard and asked for a repetition of the way in which the responses were to be announced. When these questions were answered, the experimenter proceeded to call for the judgments on the first trial. The first member of the majority was provided with the answers on a small card which he could consult inconspicuously; the others followed his lead on each trial. The experimental session lasted about 20 minutes.

Members of the majority served for considerable periods of time with successive critical subjects. At times those who had been critical subjects were invited to become members of the majority in future experiments.

---

[4] In the experiments preceding this investigation (1, pp. 455–456) the majority played a somewhat more active part. The present modification in the direction of greater passivity did not, as far as we could judge, alter the results, although it lightened the burden of the critical subjects.

[5] These instructions were adopted following a number of minor revisions with earlier subjects. The modifications of wording had no observable effect upon the results.

As mentioned earlier, the majority consisted of seven to nine persons. It was felt that this number was neither too large nor too small to produce a credible and serious sense of group contradiction. In a few instances the majority had only five or six members; this difference, we will show subsequently, did not affect the outcome. The members were not selected with any criteria in mind other than general reliability and absence of unusual features (such as visibly poor eyesight). While the composition of the group was fairly constant, some did drop out and had to be replaced, a circumstance that seemed to have no discernible effect on the results.

Both the majority and the critical subjects were asked to hold the experiment in confidence. This was a necessary condition for continued work in the same institution, and one that was achieved with a signal degree of success.

The Critical Subjects

The subjects were all male, white college students, ranging in age from 17 to 25; the mean age was 20. For certain purposes a large number of critical subjects was required for the present experiment. The present report is based on a total of 123 subjects.

The experiments were conducted in three institutions which differed in a number of respects. The first group (Group I) was drawn from a small private college of high educational standing. Its student population came largely from upper middle-class homes, from families long resident in the United States. Their parents were generally college educated, with fields of work mainly in business, technical areas, and the professions. The previous education of these students was, by current standards, superior. They had attended excellent high schools; in many cases they had received part of their education in Europe. In intelligence they were considerably above the average of the college norm. Perhaps even more pertinent to us is the social orientation of these students. Much of their previous and current discipline had the purpose of instilling in them a spirit of independence. They had been taught and had adopted the values of self-reliance, of developing personal convictions and standing up for them. The second group (Group II) came from a large metropolitan college. The population was more heterogeneous but predominantly lower middle class. A considerable proportion of the students were second-generation immigrants who had lived nearly always in an urban environment. They represented a more diversified array of professional and intellectual interests than either of the other groups. The third group (Group III) comprised students from a state teachers college, mainly from lower middle-class homes, whose educational and intellectual development was poorer than that of the other groups.

The critical subjects and their majority were always drawn from the same population. In short, we investigated the effect of a group of peers upon a minority of one. However, the degree of acquaintance with the majority was uncontrolled; it varied markedly from institution to institution, and also from person to person. In the first group the members of the majority were often acquaintances, and at times, friends of the critical subjects. This was far less frequently the case in the other groups.

The critical subjects were chosen at random from the respective populations.

The Materials

As stated earlier, the task consisted of the comparison of a standard line with three other lines, one of which was equal in length to the standard. The lengths to be compared appeared as black vertical lines on white cards that were placed on the ledge of the blackboard in front of the room. As the instructions indicated, the comparison lines were numbered 1, 2, and 3 from left to right, and the members stated their judgments by calling out one of the numbers.

The cards remained in position until all had announced their estimates; they were then removed and replaced by a new pair of cards carrying a new set of standard and comparison lines.

The lines were vertical black strips, 3/8 inches wide, pasted on white cardboards which were 17 1/2 by 6 inches. On one card appeared the standard line; the other card carried the three comparison lines. All lines started at the same level, their lower ends being 2 1/2 inches from the lower edge of the cards. The standard line appeared in the center of the card, while the comparison lines were separated by a distance of 1 3/4 inches. The comparison lines were numbered 1, 2, and 3 from left to right with black gummed figures 3/4 inches long. They were placed directly underneath the lines and 1/2 inch from their lower end. The standard and its matched comparison line were always separated by 40 inches.

The Seating Arrangement

The group distributed itself in two rows of an ordinary classroom, in which they occupied adjacent seats. As mentioned earlier, the critical subject was in the next to the last seat in the second row. The cards were placed so that the critical subject was directly between them. In Fig. 1 we present a schematic outline of the position of the group with respect to the stimulus materials.

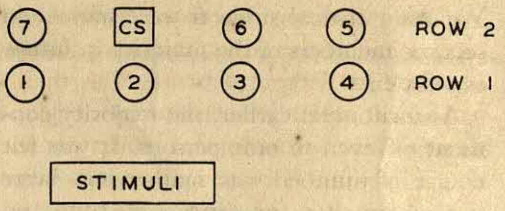


FIG. 1. Seating arrangement of majority and critical subject. The circled numbers designate the members of the majority and the order in which they announced their estimates. CS designates the critical subject.

Structure of the Task

The standard and comparison lengths, the order in which they appeared, and the responses of the majority, are included in Table 1. There was a total of 18 comparisons. The errors of the majority, which varied from 3/4 inches to 1 3/4 inches, were smallest on the early trials, generally increasing as the experiment progressed. We will now describe the task in detail.

Repetition of the series. We note first that the series consisted of nine comparisons which were shown twice without a pause.

TABLE 1

MAJORITY RESPONSES TO STANDARD AND COMPARISON LINES ON SUCCESSIVE TRIALS

Trial	Length of standard (in inches)	Length of comparison lines (in inches)			Majority error (in inches)	Type of error
a*	10	8 3/4	10	8	0	
b*	2	2	1	1 1/2	0	
1	3	**3 3/4**	4 1/4	3	+3/4	Moderate
2	5	5	**4**	6 1/2	−1	Moderate
c*	4	3	5	4	0	
3	3	3 3/4	**4 1/4**	3	+1 1/4	Extreme
4	8	6 1/4	8	**6 3/4**	−1 1/4	Moderate
5	5	5	4	**6 1/2**	+1 1/2	Extreme
6	8	**6 1/4**	8	6 3/4	−1 3/4	Extreme
d*	10	8 3/4	10	8	0	
e*	2	2	1	1 1/2	0	
7	3	**3 3/4**	4 1/4	3	+3/4	Moderate
8	5	5	**4**	6 1/2	−1	Moderate
f*	4	3	5	4	0	
9	3	3 3/4	**4 1/4**	3	+1 1/4	Extreme
10	8	6 1/4	8	**6 3/4**	−1 1/4	Moderate
11	5	5	4	**6 1/2**	+1 1/2	Extreme
12	8	**6 1/4**	8	6 3/4	−1 3/4	Extreme

* Letters of the first column designate "neutral" trials, or trials to which the majority responded correctly. The numbered trials were "critical," i.e., the majority responded incorrectly.

Bold face figures designate the incorrect majority responses.

Trials d to 12 are identical with trials a to 6; they followed each other without pause.

We may therefore speak of a first and second half of the series, the latter following the first without a break. This duplication will permit us to follow the development of the experimental effect in the course of time. The critical subjects were not aware that the series was being repeated, although occasionally some remarked about the similarity of one or another trial.

Neutral trials. It seemed advisable to include a number of trials to which the majority responded correctly; these we will call the neutral trials. We hoped that their inclusion would lend a quality of trustworthiness to the majority. For this reason we also decided to make the two opening trials neutral. The third neutral trial was interspersed in the fifth position. The same neutral trials reappeared in the second half of the series, in positions d, e, and f (see Table 1).

Critical trials. The critical trials were those to which the majority responded incorrectly. There were twelve such trials, six in each half of the series. Actually the critical trials consisted of a repetition of three basic comparisons reproduced below in Fig. 2.

Moderate and extreme critical trials. For a reason that will soon become clear we introduced a systematic and constant difference on each trial between the two unequal comparison lines. In each case one of the comparison lines deviated from the standard more than the other, and this difference was in all cases 1/2 inch.

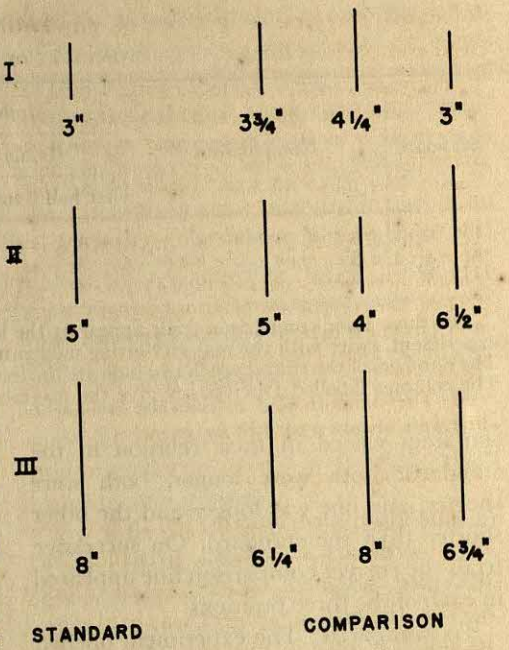


FIG. 2. Critical comparisons.

Each of the three basic comparisons was shown twice within each half of the series, but the majority responded differently to them on the two occasions. On its first appearance the majority matched the standard with the comparison line that deviated least from it; when the same lines reappeared, the majority matched the standard with the comparison line that deviated most from it. For example, Comparison I which appeared on Trials 1 and 3 called for the matching of a 3-inch standard with one of the following comparison lines: 3 3/4 inches, 4 1/4 inches, 3 inches. On Trial 1 the majority chose 3 3/4 inches as equivalent to the standard, but on Trial 3 they chose the 4 1/4 inches line. We will call the former a moderate and the latter an extreme error; and we will refer to the corresponding trials as moderate and extreme. This property of the stimulus relations will permit us to follow the effects of moderate and extreme majority errors on otherwise identical comparisons.

Table 2 exhibits the ways in which the three basic comparisons were employed in the series. It shows that each of the critical comparisons was shown four times, twice with the majority moderate and twice with the majority extreme.

The lengths of the standards varied considerably, being 3 inches, 5 inches, and 8 inches, respectively. The errors of the majority contained both over- and underestimations. The pair of unequal compari-

TABLE 2

STRUCTURE OF COMPARISONS

Standard	Comparisons	Moderate trials			Extreme trials		
		Trials		Discrepancy	Trials		Discrepancy
		1st half	2nd half		1st half	2nd half	
I 3″	3 3/4″ 4 1/4″ 3″	1	7	+3/4″	3	9	+1 1/4″
II 5″	5″ 4″ 6 1/2″	2	8	−1″	5	11	+1 1/2″
III 8″	6 1/4″ 8″ 6 3/4″	4	10	−1 1/4″	6	12	−1 3/4″

The three basic comparison trials appear at the left. Each was shown 4 times in the course of the experiment, twice with the majority erring moderately and twice with the majority erring extremely. The numbers of the trials in each row indicate the four positions in which a given comparison appeared. The columns labelled Discrepancy give the magnitude of the majority error.

son lines varied in their relation to the standard: both were longer, both were shorter, and one was longer and the other shorter than the standard. On successive trials the correct comparison line appeared in each of the three positions.

To summarize: The experiment opened with two neutral trials to which the majority responded correctly. On the third trial, the majority deviated from the correct value by 3/4 inches; on the fourth trial, the error increased to 1 inch. Another neutral trial occurred in the fifth position. The next four trials were critical, the majority error increasing progressively to 1 3/4 inches. As described above, the majority responded with moderate and extreme errors. The entire set of the first nine trials was repeated without a break.

Role of the Experimenter

The experimenter acted throughout as an "impartial chairman." He conducted the proceedings in a matter-of-fact way, reading the instructions, presenting the stimulus materials, and recording the announced estimates. When dissident judgments were given, he listened and recorded them without a show of surprise. By his behavior he indicated that he was aware of the presence of disagreement; he occasionally looked, for example, in the direction of the critical subject. But he refrained from exerting silent pressure, withholding expressions of displeasure or pleasure. His presence and example discouraged discussion and other interruptions that might have occurred in a more informal setting.

To state how the experimenter conducted himself is not, of course, to describe the effects he might have exerted. Actually he was a third force. Although he was above the clash of battle, he lent weight to the conflict. We know that subjects at times scrutinized him in order to obtain an inkling of his reactions. It is justifiable to say that the experimenter set the tone for the formal character of the session.[6] The fact that he was interested in obtaining accurate results might be considered a stimulus to independence; in addition, it may have occurred to the subject that the experimenter could well judge the accuracy of the estimates. He was also an authoritative witness, however, in front of whom the subject might find it difficult to appear a dissenter. Probably the experimenter exerted both of these effects. Although we cannot speak with assurance about his contribution, it is perhaps sufficient for the purpose of the investigation that his conduct was a constant factor.[7]

---

[6] It would be of interest to follow the course of events under more informal conditions, with the experimenter himself chosen from a group of peers or, as is also possible, without a designated experimenter.

The Postexperimental Interview

An individual interview with the subject followed immediately upon the conclusion of the experiment; the results will be described in Section IV. The interview was designed to provide information concerning the reactions to the experimental condition and to clarify the reasons for independence and its failures. In the course of the interview the purport of the experiment was explained in full. We did not consider it advisable or justified to allow subjects to leave without receiving an explanation of the procedure and of the reasons for the investigation. They were also given to understand that their performance and the interview would be held in confidence. The disclosure rarely failed to heighten interest and willingness to explore further the quality of one's reactions. The subsequent comments were as a rule more freely given, and indeed formed a valuable part of the interview.

III. THE MAJORITY EFFECT: QUANTITATIVE FINDINGS

A. *Magnitude of the Majority Effect*

The procedure just described permits a simple quantitative determination of the majority effect. An estimate of a minority subject on a critical trial may be correct and therefore independent of the majority, or it could be an error either identical with that of the majority or in its direction. (It will be made clear shortly that all errors were a function of the majority condition. See below, this page). We may therefore take the number of errors as an index of the effect the majority exerted upon a given minority of one. Since there was a total of twelve critical trials, the errors (and the independent responses) can vary from zero to twelve, inclusive. To be sure, the errors were not of the same magnitude from trial to trial; as we have seen, the trials differed in a number of respects. For the present we may postpone these distinctions and ask how often the subjects erred in relation to the erroneous judgments of the majority. The relevant results appear below in Table 3, which contains the frequency distribution of errors on critical trials of the three experimental groups.

In order to evaluate the performance of the critical subjects, we need to know how a comparable group judges when the majority condition is excluded. Accordingly the lines were judged individually by two groups, totalling 37, selected from another and comparable college population. Each subject wrote down his judgments on a prepared form, without knowledge of the estimates of his neighbors. The results of the control groups also appear in Table 3.

A reference to Table 3 reveals the following:

1. The estimates of the control group were virtually free of error. Thirty-five of 37 subjects made errorless estimates; of the remaining two subjects one showed one error, the other two errors. The proportion of errors was less than 1 per cent of the total number of critical estimates.

2. In contrast, the critical subjects showed a marked movement toward the majority. Errors increased strikingly, their frequency among individual subjects ranging from 0 to 12, or up to the maximum the conditions permitted. Only one-fourth of the subjects in the three experimental groups showed errorless performances, while in the control group 95 per cent

---

[7] The experimental sessions were conducted by a number of experimenters differing in age, sex, and other personal characteristics. We failed to find any significant differences between the performances of their subjects.

TABLE 3

Distribution of Errors in Experimental and Control Groups

Number of errors	Control group	Experimental groups			
		Group I	Group II	Group III	All experimental groups
	($N=37$)	($N=70$)	($N=25$)	($N=28$)	($N=123$)
0	35	17	5	7	29
1	1	4	2	2	8
2	1	7	1	2	10
3		12	1	4	17
4		3	1	2	6
5		5	2	0	7
6		2	4	1	7
7		3	0	1	4
8		7	4	2	13
9		3	2	1	6
10		4	1	1	6
11		2	0	2	4
12		1	2	3	6
Mean	0.08	4.01	5.16	4.71	4.41
Median	0.00	3.00	5.50	3.00	3.00
Mean per cent	0.7	33.4	43.0	39.3	36.8

were free of error. The mean number of errors in the three experimental groups varied from 4.0 to 5.2, as against a mean error of .08 in the control group. The action of the majority distorted one-third of the reported extimates, in contrast with errors of less than 1 per cent under control conditions. Inspection of the data suffices to show that the differences between each of the three experimental groups and the control group are highly significant. We also compared the frequency of subjects with and without errors (0 errors vs. 1 to 12 errors) in the combined experimental groups and the control group. The χ^2 value was 58.4, which for one degree of freedom yields a $p < .001$. (For the justification of combining the data of the three experimental groups, see p. 12.)

3. While the majority effect was considerable, it was by no means complete, or even the strongest force at work. The preponderance of estimates was, in each of the experimental groups, correct or independent of the majority, evidence that the given stimulus conditions—the facts that were being judged—were, under the circumstances, the most decisive.

These results receive a clear illustration in Fig. 3, which plots the *correct* estimates of the control and the combined experimental groups on successive critical trials. For purposes of comparison the number of correct estimates on each trial was divided by the number of subjects. (The critical trials are numbered according to their position in Table 1.)

In the control group the estimates deviate only occasionally, and then very slightly, from the level of 100 per cent accuracy, while the experimental groups show a significantly lower level of accuracy on each trial. At the same time the experimental group is much nearer to the level of the control group than to the point of

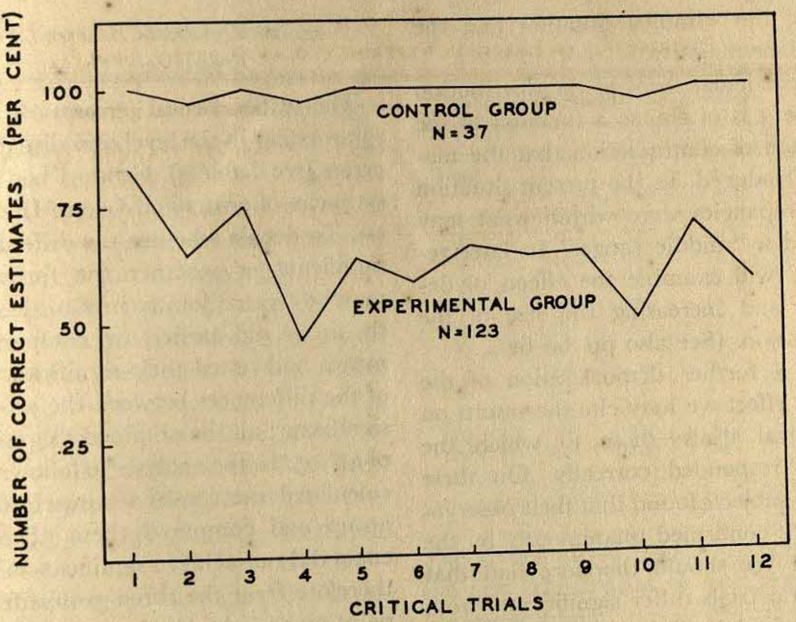


FIG. 3. Correct estimates on successive critical trials: experimental and control groups.

zero accuracy, which would represent complete domination by the majority. It is helpful to consider that the area included between the two curves represents the majority effect, while the area below the experimental curve represents the resistance to the majority.

We may also call attention to the contrast between the placid course of the control curve and the large fluctuations of the experimental curve. Under the experimental stress the critical trials display marked individual differences which they fail to show under the control conditions. Subsequently we will consider the reasons for the turbulence produced by the intervention of the majority.

4. The data of Table 3 reveal the presence of extreme individual differences in response to the experimental condition. There were completely independent subjects, and there were others who went over to the majority without exception; the distribution was continuous between these extremes. One-fourth of the experimental groups (24 per cent) gave errorless estimates, while an approximately equal number (27 per cent) gave majority-determined estimates from eight to twelve times. Between these extremes is to be found one-half of the critical group, with errors ranging from one to seven. That the majority elicited widely different reactions is one significant aspect of the present findings.

The distribution of errors departs from the normal curve often obtained in psychological measurement, being more akin to a J curve. But it differs in a fundamental respect from the J curves of conformity reported by F. H. Allport and his students. Unlike the latter, the mode occurs not at a point determined by convention or by the pressure of a group, but rather at the truth value, that is to say, at the opposite extreme from the majority position. The obtained distribution is clearly a resultant of the two major forces acting upon the subjects: the stimulus situation and the stand of the majority.

The particular form of the distribution in Table 3 is of course a function of the magnitude of contradiction that the majority introduced. In the present situation the discrepancies were within what may be called a "middle range." In later reports we will examine the effects of decreasing and increasing the size of the contradiction. (See also pp. 60–62.)

5. As a further demonstration of the majority effect we may cite the results on the neutral trials—those to which the majority responded correctly. On these trials the subjects found that their observations were confirmed unanimously by the majority. We should therefore find that the neutral trials differ significantly from the critical trials. On the other hand, we should anticipate no marked differences between these trials in the control group. The results substantiate these simple inferences. On neutral trials there were, in the experimental groups, only three errors (out of 738 judgments); the control group gave six errors (out of 222 judgments) on the same trials.

From the preceding analysis we draw the following conclusions:

1. The unanimously wrong majority produced a marked and significant distortion in the reported estimates. This is attested by the high and consistent differences between the experimental and control groups.

2. Despite the effect of the majority the preponderance of estimates was, under the present conditions, independent of the majority.

3. Individuals responded in fundamentally different ways to the opposition of the majority, ranging from complete independence to complete yielding.

B. *Differences Between the Experimental Groups*

The experimental groups differed to some extent in the level and distribution of errors (see Table 3). Group I had the lowest mean of errors and Group II the highest. To decide whether the differences are significant or whether the three groups may be treated jointly in statistical analysis, as we did earlier, we compared their means and tested their significance. None of the differences between the groups was significant; all the obtained t's gave values of $p > .05$. In the analyses to follow we have calculated the results separately for each group and compared them. The differences did not achieve significance; we will therefore treat the three groups from this point on as a single population.

The absence of clear differences between the groups, despite the social and intellectual contrasts between them, has been a source of puzzle to the writer. One might conjecture that the experiment placed the members of Group I in the most difficult position. They were most often with majorities who were their friends and acquaintances, whom they knew they would have to face subsequently; in contrast, the subjects of the other populations could disappear at the conclusion of the experiment and have no further contact with their group. And, indeed, our observation convinced us that the members of Group I reacted most deeply and were in the strongest conflict (a result that may also be related to the high standards of independence that prevailed among them). Yet we have no knowledge of the effect of intimacy on the present situation. Observations with still other and different college groups again yielded similar results. The uniformity of the obtained level of errors may reflect a widespread cultural condition that overrides the differences we have described.[8]

---

[8] There remains the possibility that the respective groups did differ in independence but that the differences were obscured by a feature of the present procedure. In each case we observed independence in the midst of a group of peers drawn from the same population as the critical subject, Conceivably the members of Group I might have found it harder to resist *their* majority than they

C. *Role of the Stimulus Conditions*

It is the presupposition of the present experiment that the stimulus conditions exert a fundamental effect on the character and course of the majority influence. This was the reason for choosing as the object of judgment facts or relations that possessed an independent status. In general, it does not make much sense to divorce the consideration of social effects from the conditions to which they refer. Group action necessarily derives its significance for the individual from the reference it has to facts, real or alleged. The crux of the present experimental situation rested precisely on the contradiction by the majority of an obvious state of affairs. It is therefore of importance to examine the dependence of the results obtained on the facts in question.

The Majority Effect as a Function of the Critical Trials

Since the critical trials differed from each other appreciably, we may now inquire whether they responded differently to the majority condition. Table 4 gives the incidence of errors on successive critical trials in each of the experimental groups. The same information appears graphically in Fig. 4; here the values are stated as percentages, to permit a direct comparison between the groups.

The frequency of errors varied considerably from trial to trial. (This we already had occasion to see in Fig. 3 which plotted the frequency of independent estimates.) The errors varied from 17.1 per cent on the first critical trial to 54.5 per cent on the fourth trial.

TABLE 4

FREQUENCY OF ERRORS ON SUCCESSIVE CRITICAL TRIALS IN THREE EXPERIMENTAL GROUPS

Trial	Group I ($N=70$)	Group II ($N=25$)	Group III ($N=28$)	All groups ($N=123$)
1	8	6	7	21
2	25	9	10	44
3	14	6	10	30
4	37	16	14	67
5	22	11	12	45
6	27	13	13	53
7	22	9	12	43
8	24	9	12	45
9	23	11	10	44
10	34	15	14	63
11	17	10	9	36
12	28	14	9	51

Before considering the possible reasons for these fluctuations, we may note their high consistency in the three experimental

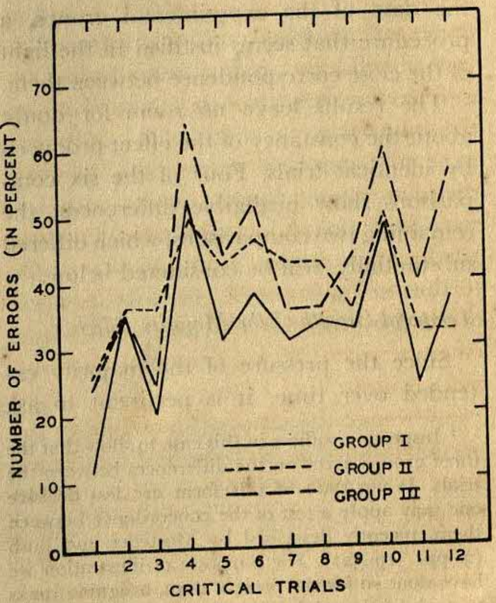


FIG. 4. Frequency of errors on successive critical trials in 3 experimental groups.

---

would those of the other groups; and the subjects from the other populations might have been less independent in the face of majorities drawn from Group I. Had our aim been to establish a generally valid index of individual independence it would have been necessary to observe subjects in settings that deliberately varied the relation between them and the opposing majorities.

groups. Trials that produced high errors in one group tended to do so in the other groups, and similarly for trials that most resisted the force of the majority. Although we have not as yet specified the basis of these regularities, we can assert that the majority effect is a lawful function of certain stimulus properties.[9] At the same time, the data offer further substantiation of the similarity between the experimental groups.

The following analysis provides further support for the conclusion that the regularity of the data is a function of the stimulus conditions. It will be recalled that the critical trials were shown twice without a break in the course of the experiment, a fact that enables us to compare the results for identical trials early and late in the series. Figure 5 provides this comparison; it plots the frequency of errors on identical trials in the first and second half of the series. For this purpose we have combined the data of the experimental groups, a procedure that seems justified in the light of the close correspondence between them.

The results leave no room for doubt about the constancy of the effect produced by identical trials. Four of the six comparisons show negligible differences; the remaining two comparisons, which differed substantially, will be considered below.

Temporal Growth of the Majority Effect

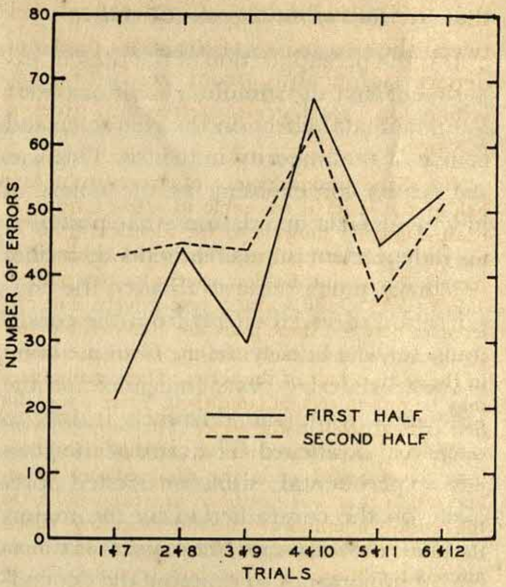


FIG. 5. Comparison of errors on identical trials.

Since the pressure of the majority extended over time, it is pertinent to ask whether the effects obtained were a function of the temporal conditions. We will now inquire whether the experimental groups came increasingly under the influence of the majority with continued exposure to it.

The relevant data have already been presented in Table 4 and in Fig. 5. They might suggest that the majority effect increased during the first four trials and reached a level from which it did not further depart. This conclusion must be regarded with reserve in view of the other differences between the trials; for example, the early rise might have disappeared if the order of the trials had been different. A somewhat clearer basis of comparison is that between the two halves of the series which were identical in all respects save temporal position (see Fig. 5). The incidence of errors is somewhat higher in the second half, the mean being 2.29 in comparison with a mean of 2.11 in the first half of the series, but the difference is slight and falls considerably short of significance ($t = 1.63$, which for 10 degrees of freedom gives a $p > .05$). In confirmation we find that the distribution of differences between the two halves of the series does not depart significantly from zero ($t=1.77$, which for 122 degrees of freedom gives a $p>.05$).

---

[9] Inspection suffices in this case to show that the three groups agree on the differences between the trials. When data of this form are less decisive one may apply a test of the concordance between them, recently described by Mosteller and Bush (3, pp. 319–321). For purposes of illustration we have done so for the present data, assigning ranks to each of the trials according to the number of errors; these ranks were assigned for each experimental group separately. We computed an approximation to chi square for the ranked data, which is equal to 64.1, which with 11 degrees of freedom is significant beyond the .01 per cent level of confidence.

The slight preponderance of errors in the first half of the series is traceable to a few particular trials. Figure 5 shows that the identical trials in the early and late halves of the series behaved in remarkably similar fashion, except for two pairs—Trials 1 and 7, and 3 and 9, respectively. The differences between these pairs are considerable and significant, and in each case the errors are larger in the second half of the series. These differences may be due to special conditions. Trial 1 was the first critical trial in the series. The subjects were completely unprepared for it. At this point they may have looked upon the disagreement as an aberration that was not likely to reappear, or as the result of a temporary misunderstanding. When the same trial reappeared, the subjects were in an altered condition; they had now been exposed to the experimental stress and had adopted a more or less consistent policy towards the majority. It is somewhat more difficult to account for the difference between Trials 3 and 9. The following is a possible explanation, although we cannot support it with independent evidence. Trial 3 was the first extreme trial encountered by the subject. and possibly the shock promoted independence that weakened with the further progress of the experiment.

The evidence fails to reveal a consistent growth of the majority effect with time. At most, it points to a limited rise during the first few trials, and this conclusion must remain indecisive in the absence of information about results with longer series and with trials more similar in construction than those here employed (see pp. 58–60). Moreover, a definite conclusion requires an examination of the performances of individual subjects, since mass temporal effects, if present, might mask differences in the rate with which different individuals began to succumb to the majority. This question will also be discussed later. (See pp. 20–21.)

Moderate and Extreme Errors

We will now describe an effect of considerable regularity that was a joint function of the stimulus conditions and of the action of the majority.

It will be recalled that the majority responded with moderate and extreme errors to identical stimulus relations. This procedure permits us to trace the dependence of moderate and extreme errors by the critical subjects upon those of the majority. The relevant data concerning the incidence of extreme and moderate errors as a function of the same errors in the majority appear in Table 5.

The data of Table 5 establish that when the majority committed a moderate error, all errors of the critical subjects were also moderate. This result is understandable and fits with earlier findings. We have seen that under control conditions there was a virtual absence of errors; these were almost completely a consequence of the majority pressure. This being the case, it would be unusual for a critical subject to

TABLE 5

ERRORS AS A FUNCTION OF MODERATE AND EXTREME MAJORITY RESPONSES

Experimental group	N	Majority moderate		Majority extreme		
		Extreme errors	Moderate errors	Extreme errors	Moderate errors	Total
I	70	0	150	98	33	131
II	25	0	64	58	7	65
III	28	0	69	53	10	63
All Groups	123	0	283	209	50	259
Per cent		0	100	80.7	19.3	100

commit an extreme error when the majority was moderate.[10]

How do erring subjects cope with the majority when it is extremist? Under this condition we find that (*a*) most errors are extreme or in accord with the majority, and that (*b*) a significant proportion of the errors is moderate. Four-fifths of the errors were identical with the majority, while one-fifth were errors intermediate in size between the majority position and the true value.[11]

The frequency of errors was very much the same on moderate and extreme trials. One may not conclude, however, that the magnitude of the contradiction is of no consequence. Under the present conditions the differences between moderate and extreme errors were small. For the identical stimulus relations the errors differed by a constant amount, namely, one-half inch. Over the entire range of comparisons the two sets of errors differed by a small amount and, indeed, overlapped. (Moderate errors had the values of 3/4 inch, 1 inch, and 1 1/4 inches, respectively; while the extreme errors were 1 1/4 inches, 1 1/2 inches, and 1 3/4 inches, respectively, as Table 2 shows.) In the present context moderate and extreme errors refer mainly to a relational property of the erroneous alternatives and not exclusively to their absolute size; a given discrepancy can be a moderate error in one constellation and an extreme error in another constellation. It is also likely that the absence of marked differences in the incidence of the two types of errors was due to the fact that they were part of the same situation. It will become clear shortly that the critical subject adopted a course of action toward the entire experimental situation which decided his over-all responses. He arrived at what is analogous to a decision to oppose or to follow the majority, which imposed a single direction upon his judgments, thus obscuring some distinctions that might otherwise emerge. It remains for further investigation to establish by means of systematic variations the effect of magnitude of discrepancy. (See pp. 60–62.)

Compromise Reactions

On the basis of the preceding evidence we propose that some errors—viz., moderate errors when the majority was extreme—had the character of compromise reactions. Being in the midst of forces proceeding from the insistent demands of reality and from the majority, the critical subjects at times chose a middle course. The following analysis supports the conclusion that moderate errors occurring when the majority was extreme did have the character of compromise reactions.

For this purpose we need to call attention to a further difference between the critical comparisons. Figure 2 and Table 2 show two kinds of relation between the unequal comparison lines and the standard. (*a*) On trials of the Form I and III both unequal comparison lines deviated from the standard in the same direction. Thus, on the I trials *both* comparison lines were *longer* than the standard; while on the III trials *both* were *shorter* than the standard: in each case, the comparison lines were in the same direction relative to the standard. (*b*) On trials of Form II the comparison lines were in opposed directions from the standard, one being shorter, the other longer than the standard.—Comparisons of the first type will be termed *relation-A* comparisons; and of the second type, *relation-B*.

---

[10] Under certain conditions a person might of course attempt to anticipate and exceed the direction of a group. This was not possible in the present situation, since the majority was not consistent in the direction of its errors from one trial to the next. One could study the operation of such a tendency by having the majority consistently under- or overestimate, sometimes moderately and sometimes extremely. In that case it would also be necessary to have the critical subject be the first to respond on each trial. This experiment was not done.

[11] Group I showed a higher proportion of moderate errors on extreme trials than either Groups II or III. The percentage of such errors was, in the three groups, 25.2, 10.8, and 10.6, respectively. This was the only occasion on which we found a marked difference among the three experimental groups.

The following difference between the two types of trials is of concern to us. On relation-A trials the moderate error lies *between* the other alternatives on the length continuum. But moderate errors of relation-B trials are so only in terms of their absolute deviation from the standard; they do not lie in the interval between the correct match and the extreme deviation. This difference becomes significant when the majority becomes extreme. If we assume, as we should, that the critical subject is acted upon by two forces, the perceived relation and the position of the majority, he is facing different problems under the two conditions. On relation-A comparisons the majority is at the extreme end of the continuum and the moderate response is in the region between the majority and the correct value. If he feels compelled to accede to both forces, he might choose the moderate error since it deviates least from both. In contrast, the moderate error of relation B lies outside the region of the two major forces; if he chooses the moderate error, the subject is removing himself as far as possible from the majority. If these considerations are valid, we should find that relation A will produce moderate errors more frequently than relation B.

The results fully support this inference, as Table 6 shows. Here we grouped the errors on majority-extreme trials according to their membership in the relation-A or relation-B category. With only a single exception moderate errors are not found on relation-B trials; they come solely from relation A. The reader will note that the present data are a breakdown of the errors appearing in the "majority extreme" columns of Table 5 above. There we found that 19 per cent of the errors under that condition were moderate. We have now traced the source of these errors to comparisons of the A form.

We conclude that a moderate error on an extreme trial should be considered a compromise response—an attempt to mediate between two opposing forces. When an alternative exists which is nearer to the truth than the majority position, but does not fulfill the condition of being also intermediate between them, it is not chosen. This result is all the more interesting when

TABLE 6

FREQUENCY OF MODERATE AND EXTREME ERRORS ON TWO TYPES OF EXTREME TRIALS

Experimental group	*N*	Relation A*		Relation B†		Total
		Extreme errors	Moderate errors	Extreme errors	Moderate errors	
I	70	59	33	39	0	131
II	25	38	6	20	1	65
III	28	32	10	21	0	63
All groups	123	129	49	80	1	259
Per cent		72.5	27.5	98.8	1.2	

* Trials 3, 6, 9, and 12.

† Trials 5 and 11.

we consider that subjects were as a rule not aware of the difference we have here described between A and B relations, and that we often had difficulty in explaining the distinction to them in the course of the subsequent interview. (Their difficulty is understandable when we consider that the comparison lines were not shown in the order of their lengths.)

Compromise Reactions and Total Performance

We have now identified several kinds of errors obtained under the present conditions. There were moderate errors obtained when the majority was moderate, extreme errors when the majority was extreme, and there were moderate errors with the majority extreme. The latter (which occurred almost exclusively on relation-A trials) constitute the set of compromise reactions. We will now inquire into the relations between the compromise reactions and the quality of total performance. In particular, it occurred to us to ask whether those with compromise reactions were also prone to be intermediate in their total performance, or to fluctuate between independence and submission. If a relation were present between these measures, we would be substantiating further the compromise quality of each; at the same time we would be taking a step toward establishing a quality of personal consistency in the performances.

A word of explanation is needed about the calculation. It would not be advisable to compare a person's compromise errors with his total errors, since the latter figure includes the former. Nor is it entirely adequate to compare the frequency of compromise errors with that of all other errors, since the occurrence of the former limits the occurrence of the latter. For example, a subject with three compromise errors cannot show more than nine other errors. We therefore compared compromise errors with the ratio of other errors to the maximum available to the subject. For example, if a subject showed two compromise errors and six other errors, the latter would be expressed by the ratio 6:10, or 60 per cent.

TABLE 7

RELATION OF COMPROMISE ERRORS TO OTHER ERRORS

Number of compromise errors	Number of majority responses on remaining trials (*in percentages*)		
	0%	5–75%	76–100%
No compromise errors	29	31	27
One or more compromise errors	5	27	3

Note.—Compromise errors are moderate errors occurring on those majority-extreme trials in which the moderate alternative lies on the continuum between the true value and the position of the majority. The values above are therefore based on Trials 3, 6, 9, and 12 (see Table 6).

The relevant data appear in Table 7. We divided the distribution of all errors other than compromise errors into three parts. The first and last subgroups contained approximately 25 per cent of the cases each, and the middle class approximately 50 per cent of the cases. The results definitely support the suspected relation. Of the 34 subjects without errors on the remaining trials, only five (or 15 per cent) gave one or more moderate or compromise errors. Similarly, only three of 30 subjects who followed the majority predominantly gave moderate or compromise errors. The association between the two measures is highly significant, the χ^2 value being 18, which for 2 degrees of freedom gives a $p \overline{<} .01$.

However, the subjects who were intermediate in independence were not preponderantly those who gave moderate errors. That the converse of the main conclusion does not hold is partly due to the fact that the number of subjects with compromise errors was relatively small, totaling 35.

Role of Individual Trials

The individual trials differed from one another markedly and consistently in their susceptibility to the majority effect (see pp. 13–14.) What properties

TABLE 8

Errors on Different Types of Critical Trials

Trial	Number of errors	Absolute discrepancy (in inches)	Moderate or extreme	Size of standard (in inches)
1	21	3/4	M	3
3	30	1 1/4	E	3
11	36	1 1/2	E	5
7	43	3/4	M	3
2	44	1	M	5
9	44	1 1/4	E	3
8	45	1	M	5
5	45	1 1/2	E	5
12	51	1 3/4	E	8
6	53	1 3/4	E	8
10	63	1 1/4	M	8
4	67	1 1/4	M	8

of the stimuli were responsible for these differences? In Table 8 we have ordered the trials according to *increasing number of errors*, and have included, for each trial, the size of the majority discrepancy, the type of majority error (moderate or extreme), and the size of standard.

(*a*) We find no correspondence between the size of the majority error and incidence of errors. In general, the most error-producing trials are also those involving large discrepancies, but trials with low discrepancies are not freest of error. (*b*) Nor is there any discernible relation between relative discrepancy and frequency of errors (defining relative discrepancy as the ratio of the majority error to the size of standard.) (*c*) Also, moderate and extreme trials do not differentiate themselves (as we have already noted earlier; see p. 16). (*d*) And, as has been pointed out, there is no relation between errors and temporal position. (*e*) Evidence for a positive relation does emerge when we take the size of the standard as the basis. Save for two reversals there is a consistent positive relation between size of standard and frequency of errors. In Table 9 we divided the critical trials into three groups according to length of standard and computed the errors in each group. (The stimulus comparisons in each group were identical, as reference to Table 2 will show.) There is a progressive increase of errors with length of standard. For a subsequent confirmation of this finding see p. 61.

TABLE 9

Relation of Errors to Size of Standard

	Size of standard		
	3″	5″	8″
Trials	1, 3, 7, 9	2, 5, 8, 11	4, 6, 10, 12
Number of errors	138	170	234

The negative findings reported in this section need to be viewed with reserve. The present procedure, because it simultaneously varied many of the relevant factors, was not designed to exhibit systematically the role of stimulus-relations. It would therefore be unwarranted to conclude that a factor such as magnitude of majority error plays no role. (See pp. 60–62.)

D. *Personal Consistency*

Under the identical experimental conditions the performance of different individuals ranged from complete independence to complete yielding (See Table 3). It follows that of the various conditions that determine the reactions to the majority none are more decisive than those that reside in the person himself. To reach an adequate understanding of the majority effect, we must therefore take into account the personal factor. Here we will consider one, relatively limited, phase of this problem.

We ask whether individuals were consistent in performance. In the following we shall rely on an internal analysis of the consistency of subjects within the same experimental setting.

Internal Consistency of Individual Performance

As one main source of evidence we shall employ the performances of the same individuals in the two halves of the series which consisted of identical sets of trials. This analysis is simplified by the fact, established earlier, that the two halves did not differ significantly for the population as a whole (see Table 4 and Fig. 4.) We now compared the performances of individuals with the aim of deciding whether those who were most independent in the early part of the experiment continued to be independent, and conversely.

There is a striking relation between the

TABLE 10

Errors on First and Second Halves of Experiment

		Second half	
		0–2 errors	3–6 errors
First half	0–2 errors	63	12
	3–6 errors	10	38

performances in the two parts; those who were independent early in the experimental session continued independent, and similarly for those who yielded.[12] In Table 10 we dichotomized the scores into 0 to 2 and 3 to 6 errors, respectively. The association is significant, the χ^2 value being 48.4, which for one degree of freedom gives a $p < .01$. With the data dichotomized as shown in Table 10 we also obtained a Sheppard "U" coefficient of +.84 which is significant by a chi-square test beyond a level of confidence of .01.

We also divided the subjects according to the number of errors (ranging from 0 to 6, respectively) in the first half of the experiment, and for each group of subjects computed the average number of errors in the second half of the experiment. These were: 0.33, 1.2, 2.6, 2.8, 4.0, 4.2, and 5.7, respectively.

The foregoing analysis establishes conclusively that the performances of individual subjects were highly consistent. An individual who sided with the majority in the early trials can be predicted with a great deal of accuracy to persist in yielding; one who started as independent was highly likely to continue independent.

The interpretation of this finding however, is not simple. Consistency under the given conditions could be the result of two distinct psychological causes. (*a*) It could be the expression of a characteristic of the person—of a capacity to remain independent in action and to retain confidence in one's experience against the opposition of others. It was not within the scope of the present investigation to establish whether such a genuinely personal property was responsible for resistance to the majority effect. For this purpose it would be necessary to observe and compare the performances of the same individual in several situations each of which requires a choice between independence and submission. (*b*) In the absence of such evidence, another interpretation must also be considered as a definite possibility, namely, that the consistency observed is specific to the situation studied and may not extend beyond it. A person entering the situation may, for quite contingent reasons, grasp it in a particular way which leads him to yield or to resist (or to be intermediate) in the early trials. Whatever the reasons for his early direction, he might persist in it because it is a direction and because his first actions have committed him to a course. (Evidence from subsequent experiments shows that a factor of commitment can exert a decisive effect on future action.) The results fail to decide between these alternatives.

Latency of the Majority Effect

The majority exerted pressure during a succession of trials extending in time. It is therefore pertinent to ask how early or how late individuals gave evidence of the pressure of the majority.

Certain facts relevant to this question have been presented earlier. We know that there were completely independent and completely yielding individuals, or that some went with the majority on the first possible occasion and others not at all. Most subjects did not belong to either of these extremes; we will now inquire concerning the span of their resistance to the majority.

We ask whether the point of occurrence of *first* errors has a describable regularity. Since the completely independent individuals are for this purpose a special group,

---

[12] Contributing to the correlation is the behavior of the extreme subjects at both ends of the distribution (see Table 3). In particular, there is the large group of 29 individuals who were independent in each half. A question might be raised about the justification for their inclusion in the present analysis. It might be contended that these subjects have a special position, and that to include them bolsters the relation artificially. The argument is fallacious, since it assumes without warrant a knowledge of the total performance of a subject before he is included in a correlation chart. It is by no means obvious that a person who is completely independent in the early part of the experiment will continue to be so; this is precisely the relation we wish to clarify.

TABLE 11

POINT OF FIRST YIELDING AND RELATION TO FREQUENCY OF SUBSEQUENT ERRORS

Position of first error on trial	Number of errors following first error												N	Mean number of errors subsequent to first error	Adjusted mean number of errors subsequent to first error
	0	1	2	3	4	5	6	7	8	9	10	11			
1		1	2			1	1	5	3	1	1	6	21	7.6	7.6
2		1	3	3	4	3	2	6	3	5	3		33	6.1	6.7
3	1		1	1		1	1	2					7	4.3	5.2
4	2	2	8	2	3	1							18	2.3	3.6
5		2	2			1							5	2.2	3.4
6	2	1											3	.3	.6
7	1	1											2	.5	1.1
8		1	1										2	1.5	4.1
9		1											1	1.0	3.6
10															
11															
12	2												2	.00	

the analysis will be based on those subjects (94 in all) who went with the majority at least once.

Table 11 contains the detailed data concerning the locus of first errors. By far the greatest number of these errors is to be found in the very early trials, the mode being at the second critical trial. Following a drop at the third trial, the rate of first errors again rises on the fourth. Of those who erred at all, over 50 per cent did so by the second critical trial. Only 7 per cent erred for the first time in the second half of the series. The regularity with which first errors fall off is impressive when one considers that the successive trials differed in many regards.

A cumulative percentage curve of first errors (with Trials 1 to 12 listed on the X axis) gives a steeply rising ogive which bends, after Trial 4, to approach gradually the maximum of 76 per cent. (The calculation of percentage of error was based on an *N* of 123; it will be recalled (see Table 3) that 24 per cent of the experimental group was completely independent.)

Those who followed the majority did so predominantly in the early trials. Very few who withstood the majority at the outset weakened subsequently, an outstanding instance being the group of 29 completely independent subjects. This result could not have been readily anticipated; it would have been equally reasonable to infer that with increasing exposure more and more individuals would come under the sway of the majority. We conclude that generally the subject responds quickly to the experimental stress by adopting a consistent course of action, which is not altered substantially by the further pressure of the majority. (For the effect of a more prolonged exercise of pressure by the majority, see Section V, Experiment 5, pp. 58–60.)

The Relation of Latency and Independence

Although first errors were concentrated at the early trials, there were still considerable individual differences, some turning to the majority earlier than others. This fact permits us to explore a further question bearing on the problem of personal consistency. Is there a relation between the locus of the first error and the total majority effect for a given subject?

The answer requires a comparison of the locus of first errors with the number of

Bureau Ednl. Psy. Research
DAVID HARE TRAINING COLLEGE

errors on subsequent trials. In Table 11 we have included the relevant data, calculating for each subgroup the mean number of errors after the first error. We see that the frequency of errors is a nicely decreasing function of the locus of first errors. Those who erred on the earliest critical trial (Trial 1) show a mean error score of 7.6 on the following eleven critical trials; when first yielding occurred on the second critical trial, the mean score drops to 6.1; if on the third critical trial, the score drops still further to 4.3, and so on.

Before we draw the indicated conclusion it is necessary to take into account and correct for an obvious source of error. Clearly the mean number of errors must become progressively less as the first error occurs later and later in the series. A person who errs on the first critical trial can err (or not) on eleven subsequent trials, whereas one who first yields on Trial 9 cannot err subsequently more than three times. The needed adjustment is found by multiplying the values in the second row of Table 11 by the ratio $11/k$, where k is the number of critical trials subsequent to the first error. By this means we are in effect stating the number of errors which each subject would have shown if eleven critical trials had followed his point of first yielding. The adjusted means are given in the terminal column of Table 11.

The adjusted values parallel the uncorrected values quite closely. There is an impressive drop in the number of errors as the first error appears later in the series. The decelerating trend is compelling for the first half of the series, but it is followed by an abrupt rise on the last three trials. This reversal is rather less important than its appearance would suggest. It is based on a negligibly small number of cases, there being only five subjects following Trial 6.

A further calculation, which avoids the difficulties encountered above, fully substantiates the conclusion just reached. We have categorized the subjects according to the locus of first errors in the first half of the series and computed the errors of each subgroup in the second half of the series. Since all subjects have the same alternatives available to them in the second half of the series regardless of their performance in the first half of the series, the comparison does not require adjustment. The results, which appear in Table 11-a, again show a strong relation between the rate of first errors and the magnitude of later errors.[13]

We have shown that the span of a person's endurance against the majority was significantly related to his subsequent independence. This result confirms and throws further light upon the regularities reported earlier. Consider, for example, the case of intermediate subjects. Of them we can now say that they alternated between independence and yielding, that they chose compromise reactions more often than others (see Table 7), and that they were also intermediate in rate of yielding (see Table 11). In short, intermediacy, too, was a consistent form of action. More generally, the present result clarifies the previous findings concerning the substantial similarity between the early

---

[13] When considering the question of individual consistency it is necessary to bear in mind that the difficulty posed by the majority arose unexpectedly and continued virtually without interruption until the conclusion of the session. It might be contended that the subjects were caught off guard, that the absence of a "breathing spell" prevented them from reconsidering what they were doing and from recovering their balance. This consideration suggests the advisability of an experimental variation in which a "pause" is introduced in the middle of the session. The pause might be devoted to a peripheral and simple task, one that would provide time for reflection upon the preceding episode; or it might actually call upon the subject to answer certain questions concerning the immediately preceding events.

TABLE 11-a

POINT OF FIRST YIELDING IN FIRST HALF OF SERIES AND RELATION TO FREQUENCY OF ERRORS IN SECOND HALF OF SERIES

Measure	Trial						
	1	2	3	4	5	6	6+*
Number of subjects who first erred on a given trial in first half of series	21	33	7	18	5	3	36
Mean number of errors in second half of series	4.24	3.61	2.57	1.83	2.00	0.33	0.33

* This column refers to the subgroup of 36 subjects who were completely independent in the first half of the experimental session.

and late halves of the series (see Table 10) and concerning the lack of a consistent temporal cumulation of the majority effect (see pp. 14–15). Basically these findings grow out of the fact that early independence conferred immunity from group pressure, and early yielding perpetuated itself.[14]

The Problem of Cumulative Majority Pressure

The finding that some subjects followed the majority much earlier than others requires us to reconsider an earlier conclusion that the effect of the majority was not cumulative (see p. 14). It would now appear that some were able to retain independence for a number of trials but weakened with further exposure to the majority, a change that might be accounted for in terms of mounting pressure. This interpretation would have been quite convincing had we also found that more and more (previously independent) subjects began to err as the experiment continued. That this was *not* the rule, that relatively *few* surrendered independence on the later trials, argues against this interpretation. But the argument is not decisive. We will have occasion to show in subsequent reports that under the present conditions subjects tend to adhere to a course of action once adopted because they feel committed to maintain their consistency. It may therefore be that the failure to observe a stronger temporal effect is due to the formation of a specific counter force, that of public commitment. A somewhat more serious objection derives from the finding just reported that those who began to err relatively late subsequently erred less. Even this finding does not constitute a decisive objection. We might conjecture that those who once tasted independence had greater possibilities for coping with the mounting pressure.

In the light of these considerations we must reserve a final conclusion. The force of the majority may have been strong enough at the very outset to obscure, in the case of most subjects, a cumulative effect; and the presence of contrary forces may have further served to disguise its presence. To arrive at a more definite conclusion it may be necessary to observe the continuous operation of a *weaker* majority challenge (which might be produced by reducing substantially the magnitude of contradiction), and by attempts to eliminate the factor of public commitment.

E. Summary of Quantitative Findings

1. The procedure of creating a public disagreement between a unanimous and

---

[14] There remains the possibility that those who were entirely or preponderantly independent would have joined the majority if the experimental series were prolonged. For a further study of this question see pp. 58–60.

wrong majority and a minority of one about a concrete and simple fact produced a distortion of considerable magnitude. (*a*) Whereas individuals normally judged the relations in question with almost complete accuracy, the majority succeeded in deflecting one-third of the minority estimates in its direction. (*b*) Despite this large effect, the preponderance of judgments was independent, evidence that under the present conditions the force of the perceived data far exceeded that of the majority.

2. There were great individual differences in response to the conflict between the individual's judgment and the united contradiction of the majority. Individuals ranged from complete independence of the majority to complete submission to it.

3. The detailed findings demonstrated a far-reaching lawfulness of the majority effect. It was shown that the reactions to the experimental episode were a function of (*a*) the stimulus conditions that were the object of judgment, (*b*) the judgments of the majority, and (*c*) of the subject himself.

4. Several lines of evidence converge to show that the majority effect was a function of the stimulus conditions. (*a*) The size of the majority effect was similar in three experimental groups, and the more detailed reactions were also substantially alike. (*b*) Identical stimulus comparisons produced consistently similar effects. (*c*) Certain kinds of errors, such as compromise reactions, were unambiguously related to particular stimulus constellations. (*d*) The frequency of errors increased with the size of the standard.

5. The results also varied directly with the action of the majority. A moderately erroneous majority produced only moderate errors. When the majority was extremist, a significant proportion of the errors were compromise reactions.

6. The performances of individuals showed a high internal consistency. (*a*) Individuals showed a strong tendency to be consistently independent, yielding, or intermediate throughout the experimental episode. There was a high association between the individual's performance during the early and late halves of the experimental situation. (*b*) Compromise errors were found most frequently among individuals who were also intermediate in independence. (*c*) There was a close relation between the latency of reaction to the majority and the level of independence. Those who sided with the majority earliest also did so most frequently.

IV. The Majority Effect: Qualitative Observations

The foregoing account of the quantitative findings omitted much—indeed most—of what took place in the experimental setting. It deliberately abstracted from the circumstances that invested the situation of the minority person with significance, from his doubts and sufferings, his strengths and weaknesses. In this section we will undertake to describe the reactions to the experimental problem and the different ways in which the subjects coped with it. For evidence we will rely mainly on their reports in the postexperimental interview, and also on our observations.

This account will be limited by the shortcomings inherent in qualitative observations of complex happenings. Although we observed persons in a brief and specific setting, their reactions were too many-sided, and our means of observation too blunt, to provide more than a partial picture. Much of relevance doubtless escaped our notice or was concealed by the subjects. Further, in the absence of exact conceptual categories, we have no choice but to proceed in a descriptive way that necessarily lacks the sharpness of quantitative statements. Yet there can be little question of the need for a characterization of the emotions and ideas which the experimental setting provoked.

The account falls into three parts. We first reproduce the main parts of the interview, the replies to which form the data to be examined in this section. Second, we will describe the main phases of the reactions to the experimental episode. In this part we will stress those features that were shared by most, dealing secondarily with the individual differences. In the third part we will attempt to describe the major individual modes of reaction to the entire experimental setting.

A. The Postexperimental Interview

The interview followed directly upon the conclusion of the experimental session. During the greater part of the interview the subject was still under the stress that the experimental conditions had created. As it neared the end, the object of the experiment was fully explained and the reactions to the disclosure were obtained.

Below we reproduce the major portion of the interview, upon which the subsequent discussion depends. (The writer will be glad to supply copies of the complete interview to interested readers.)

1. Would you describe in your own words your experiences during this experiment?

2a. When you gave an estimate that disagreed with the others, did you feel that if the lines were measured with a ruler you would turn out to be right or wrong? (Right——, Wrong——)

b. Then did you feel that if the lines were measured, the others would turn out to be wrong? (Right——, Wrong——)

3. Would you say that you were concerned about the disagreements? (Yes——, No——, ?——)

4a. Still thinking back to the time that you were giving your judgments, would you say that the others made you doubtful about your accuracy? (Yes——, No——, ?——)

b. (If answer to [a] is yes):
What was the nature of the doubt? Try to describe it.

c. Did you have a great deal of doubt? (Yes——, No——)
[If no]: Then, would you say that your doubt was moderate or slight? (Moderate——, Slight——)

5a. Would you say that you were tempted at times to answer as the others did? (Yes——, No——, ?——)

b. Would you say this temptation was strong, considerable, moderate, or slight? (Strong——, Considerable——, Moderate——, Slight——)

c. [If yes]: Try to describe the reasons as carefully as you can.

The questions that followed were designed to cast light on the subject's awareness of his errors and on his reactions to the majority.

6a. Did you ever answer as the others did, against your own first choice? (Yes——, No——, ?——)

b. (If answer to [a] is yes): How often do you remember doing so?

c. (If answer to [a] is yes): On these occasions, as you continued to look at a pair of cards, did the answers of the others continue to look wrong, or did they begin to look more right? (Right——, Wrong——, ?——)

7a. What did you think about the other people in the group when they all gave an answer that looked wrong to you?

b. Did you wonder what everybody might be thinking of you when you disagreed? (Yes——, No——, ?——)

8. While you were comparing the lines, what did you think our purpose was in this experiment?

The preceding questions were put to the subject before he received any explanation of the procedure or of the soundness of his estimates. At this point we disclosed the purpose of the experiment. This was done in two steps. The first step, which was couched in hypothetical form, was worded as follows:

> Suppose I told you that your judgments were always right when you disagreed with the group, what would you think?[16]

Having obtained the answer to the hypothetical question, we proceeded to the full disclosure as follows:

> I now want to explain to you fully the character of this experiment. But before doing so, I would

---

[16] Obviously this question was not asked of subjects who went with the majority throughout.

like you to promise me to keep what I am about to tell you in the strictest confidence. When you have heard about the experiment, you will understand why. In return, you have our promise to keep in confidence anything you tell us during this interview. This was not primarily an experiment in optical discrimination. The aim was to throw light on the effect of groups on individuals in their midst. The group had been instructed in advance to give from time to time answers that were wrong. You were the only person who was not aware of this fact. The object of this investigation is to understand more clearly how people react to such conditions.

You now understand the purpose of this experiment. Because you have been through it, any further information you can give us as to what you have experienced will be of great value to us. You are now no longer a subject, but a person who possesses some valuable information about the experiment. Have you any further comments?

At this point we introduced the following question concerning any possible suspicion the subject may have felt.

a. Did you suspect that the group intended to affect your judgment? (Yes____, No____, ?____)
b. When did you first begin to feel suspicious?
c. Did you retain this suspicion?

Subjects who definitely suspected the purport of the experiment were eliminated. In most cases their behavior prior to these questions indicated clearly whether they had definite suspicion. We speak of "definite" suspicion in contrast to suspicion which a number of subjects entertained as a temporary hypothesis at some point during the experimental session; the latter subjects were not eliminated.[16] (For a further discussion of suspicion see p. 29.)

The final step consisted of a confrontation of subjects who had erred six times or more with several of the comparisons on which they had yielded to the majority. We asked the subject to match the lines again, informed him of his previous estimate and invited him to explain his reasons for it. The questions were as follows:

Let us look at these lines again.

a. What is the answer here?
b. Do you remember what your answer was previously?
c. Do you remember what the group answer was?
d. The answer you gave previously was ________; this was also the answer of the group. Could you help us to understand why you followed the group in this case?
e. Did you think this answer was correct at the time you gave it? (Yes____, No____, ?____)
f. Did you ever give an answer which you knew to be wrong? (Yes____, No____, ?____)
g. Did the differences between the lines appear less at that time than they do now?
h. (If answer to [f] was yes): How did you feel when you saw the lines one way and answered in another way?
i. (If [h] was asked): When you gave such an answer, did you at times forget about the question of accuracy, and consider rather that it might make you appear different from the others? (Yes____, No____, ?____)

How many people in this group do you know?____
How many do you know well?____

The interview contained some questions to which there were standard answers and others that called for free replies. Answers to the more objective questions were analyzed quantitatively. In addition we will draw upon the verbatim statements of the subjects for concrete illustrations of their reactions; these were taken down by the interviewer as the subject was speaking.

Before proceeding, certain limitations of the interview data should be mentioned:

1. In the course of experimentation the interview form was modified repeatedly as new data revealed further avenues of investigation. This was considered advisable since our object was to understand more fully the behavior of the subjects. But this procedure also reduced the size

---

[16] One might wonder whether these subjects did not feel justified in erring once they began to doubt the majority. Some of the protocols say just this (see p. 32). Our observations and their other comments convince us that this was not the case. We are rather inclined to hold that the conjunction of suspicion and yielding is added evidence of the power of the majority, and that the mere hypothesis of suspicion was too weak to liberate the individual. The attempt of some to explain their action in this manner is one instance in which it seems justified to speak of rationalization.

of the population for purposes of formal analysis of particular questions. Further, in order to keep the formality of the interview at the minimum, the experimenter omitted questions when they did not seem meaningful at the time to the subject, thus further reducing the number of cases.

2. The subjects differed in understanding and articulateness. Often they were unable to answer the questions or did so only perfunctorily.

3. Since the greater part of the interview was conducted before the subjects had received an explanation of the procedure, the answers were often guarded and not particularly frank. Those who had erred frequently had particular difficulty in answering many of the questions, and this was also, although in lesser measure, true of independent subjects who were wondering why they had been selected for questioning. Further, the subjects were often confused; they had not had the opportunity to sort out their thoughts and to consider quietly what had happened. It is therefore not surprising to find evidences of inconsistency in some of their statements and of outright contradiction in others. For example, it was by no means unusual for subjects to minimize at the outset their doubts and temptations, only to acknowledge them openly after they had received an explanation of the purport of the procedure. It is safe to say that the responses as a whole minimized the effect of the majority.

B. Reactions to the Experimental Conditions

First Reactions to the Contradiction

All subjects noted the disagreement as soon as it occurred. They were also aware that the majority was unanimous and that they were in a minority of one. Their first reaction was one of puzzlement. The subjects felt that there was something wrong, but could not locate the source of the difficulty. Numerous statements testify to this. At this early point they still believed that the disagreement was temporary.

As the disagreement persisted, the subjects tried to form some explanatory hypotheses. These were of many different kinds. Some wondered whether they misunderstood the instructions. "I thought they were measuring width after a while" [5].[17] Others spoke more or less vaguely in terms of optical illusions. "Thought there was some trick to it—optical illusion" [8]. "I knew I was being fooled by the lines being together and tried to allow for the illusion" [3]. Still others asked themselves whether their position or the tilt of their head might be responsible. One of the most interesting proposals, which usually came later in time, was that the group followed the first person who, for some reason, was estimating wrongly. The hypotheses had the obvious function of resolving the paradox and of mitigating the growth of concern. But they were adopted halfheartedly and were readily relinquished as the disagreement recurred.

Contributing to the difficulty of reaching a settled interpretation and enhancing the sense of uncertainty was the unpredictability of the majority from trial to trial. Without apparent cause the majority estimated correctly (on the neutral trials), awakening in some the hope that the painful disagreement would soon come to an end, and deepening the confusion of others. Nor did the majority follow any system in its misjudgments. Now it unanimously overestimated, only to underestimate the next moment; at one point it erred moderately, at another extremely. It was our impression that the fluctuations of the majority further unsettled the subjects and subtracted from their belief in the hypotheses they formed.[18]

---

[17] The numbers appearing in brackets following the quoted statements of a subject refer to the number of his errors in the experimental situation.

The Growth of Concern, Doubt, and Temptation

The further continuation of the cleavage heightened tension and aroused the more serious suspicion that either the subject or the majority was judging wrongly. The earliest hypotheses, such as those mentioned in the preceding paragraph, started from the presupposition that all were estimating efficiently and assigned the cause of disagreement mainly to objective factors external to the judgment process proper—position with respect to the stimuli, the property of the stimuli being judged, and so on. These were now abandoned and replaced by the question: Who is right, and who is wrong? This question received radically different answers, but most subjects, including the staunchest independents, at some time felt doubt about their accuracy, while the most pliable subjects at times felt the majority to be wrong.

Some of the reasons for the growing self-doubts had an apparently objective cast: "I was beginning to think that '140 million Frenchmen can't be wrong,' which is hypocritical, of course"[3]; "There's a greater probability of eight being right" [2]; "They must have been objectively correct if eight out of nine disagreed" [4]; "After all, the majority rules, so I guess I was wrong" [10]. In time, however, the subjects became more explicitly concerned about themselves, wondering whether their eyesight or judgment was misleading them. (As far as we could tell, no subjects felt concern about the group even when convinced that the latter were misjudging.) "I thought that maybe because I wore glasses there was some defect" [10]. "At first I thought I had the wrong instructions, then that something was wrong with my eyes and my head" [3]. "Maybe my eyes were going bad" [6]. "My whole mental processes were working abnormally" [6]. "Maybe something's the matter with me, either mentally or physically" [12].

The subjects were unable, however, to maintain their doubt as the clarity of the facts continued to loom before them. This lent their doubt a quality of fluctuation. "First I thought something was the matter with me or most of them" [0]. "I was sure they were wrong, but not sure I was right" [3]—a statement that nicely illustrates the hold that group opinion can have. Going in the same direction was the following statement of a subject who, after exclaiming, "Everybody here's crazy but me," added: "Seeing they had the power of numbers, I thought they must be right somehow, but it wasn't what I saw, and I think I was right" [0]. "Either these guys were crazy or I was—I hadn't made up my mind which. . . . I was wondering if my judgments really were as poor as they seemed to be, but at the same time I had the feeling that I was seeing them right. It was a conflicting situation" [8].

There were considerable differences between independent and yielding subjects in the inferences they drew from their doubt. There were at the one extreme such abject statements as: "I know the group can't be wrong" [8]; "I felt they must be right and I must be wrong" [5]. Some yielding subjects supposed that the majority were subject to an illusion, and felt their own failure to have the illusion to be a sign of defect: "Perhaps it was an optical illusion which the others had grasped and I hadn't . . . at that point it seemed defective not to have the illusion they had" [9]. At the other extreme we find comments such as the following: "I'd swear I was right—the only way to prove it to me is to measure them" [1]. "I absolutely feel I was right; I think they were absolutely wrong" [0]. "I felt sincerely I was right; I would stick by my decision until I was proven wrong" [0]. "I wouldn't have said what I did if I didn't think I was right" [1]. It seemed to the writer that one characteristic of many independent subjects was not so much their immunity from doubt as their ability to free themselves of it, of which the following statements are an example: "It could have been my error, but when it happened again and again I knew they were wrong" [1]. "At first I felt there was something wrong with me, that there was something about my make-up that caused me to be different, then I gained more confidence and thought I might be as right as they" [0]. "Yes, they did arouse doubt, but after I thought about it I still felt I was right" [0].

Only rarely did we find an independent subject completely free of doubt. Comments such as the following were quite unusual: "Gee, I feel like

---

[18] It would be of definite interest to vary systematically the predictability of the erring opposition. If, for example, the majority uniformly over- or underestimated (or if the majority employed a transparent system, such as always selecting the comparison line in the middle), there would be new alternatives open to the critical subjects as they encompassed the situation in a general formulation. At the other extreme it would be feasible to arrange for the majority to be visibly inconsistent (e.g., by judging clearly identifiable stimuli in contradictory ways on successive occasions).

Columbus—I feel the others were wrong. I wondered how they could possibly see something different from what I saw" [o]. But even this subject was not free of a kind of doubt, as became apparent in his reaction to the disclosure. In response to the first step of disclosure (see p. 25), he stated: "I would almost expect that, to be honest with you,—but if you tell me they weren't correct, by God, I wouldn't know what to think. Then I would have to believe that what is right is wrong." Then, following upon the full disclosure, he stated: "I thought so, but wondered if I had paranoid tendencies. I felt a bit flushed, but I answered as I saw them. Perhaps if there were a hundred people in there I might have gone along."

Many independent subjects were reluctant to question the accuracy of the majority even when they were confident of their own judgments. (Would you say that the group judged wrongly? "It's difficult to put it that way; let's say I saw them that way.") Viewed externally, it may appear that they were inconsistent or confused, but this conclusion is at variance with our observations. Their attempt to skirt the question was due mainly to the absence of proof and to the advantage the majority had in terms of sheer probability.

In order to escape the dilemma some independent subjects retreated to a phenomenalistic position, restricting themselves to the assertion that they were reporting what they saw. On this point at least their position brooked no contradiction.[19] "Hesitate to say that I was right. All I know is what I saw, and I think I may be right" [1]. "That's the way it looked to me" [2].

The reader may wonder how it was possible to maintain the credibility of the majority without arousing suspicion. Actually instances of suspicion were rather infrequent, a finding that has an uncomplicated basis. In the first place the subjects assumed that the majority was as sober and concerned as they were; the setting of a formal experiment powerfully contributed to this presumption. Nevertheless, one would expect distrust to grow as the majority continued to err. It seems to us that the subject's quick-growing concern about himself forestalled this development. Before his suspicions had the opportunity to take root, he had unwittingly come to doubt himself and to seek for explanations, as we have seen, in other directions. As a result, when subsequently he did entertain suspicion, the thought lacked decisive force, and appeared to be a case of special pleading.

Indeed, it crossed the minds of many subjects in the course of the experiment that the majority might be deliberately misleading, or that the group was following the first member who, for some unknown reason, was in error. However, this belief had the form of a fleeting hypothesis which, like many others, came and went without altering the course of action. The best evidence for this is that such expressions did not prevent subjects from yielding heavily. Actually we had no difficulty in distinguishing between full and hypothetical suspicion. In the former case, the demeanor of the subject altered completely. He might loudly inform the majority of his conclusion, be highly amused, and cease to treat their estimates seriously. In the latter case the subject remained under the stress of the forces produced by the conditions.

We shall now report some of the quantitative findings in response to the standardized questions. The responses were studied as a function of the performances of the subjects. For purposes of comparison we dichotomized the population into those with 0 to 2 errors (which was the range of errors found under control conditions) and with 3 to 12 errors, respectively. In the present discussion we refer to the former group as independent and to the latter as yielding.

---

[19] The reports of subjects that they at times went with the majority because they did not wish to appear in error suggested to Dr. Dorothy Dinnerstein a procedure that deserves investigation. The task could be altered to require the subject to report, not an objective judgment of relative length, but the way the lengths appear to him. The subject's task would then become strictly that of reporting a phenomenal datum. Provided they could undertstand and adhere to the instructions, the procedure would furnish an index of willingness to report a phenomenal datum at variance with that of others.

Ninety-four subjects answered the following question: "When you gave an estimate that disagreed with the others, did you feel that if the lines were measured with a ruler you would turn out to be right or wrong?"[20] Of these, 67 subjects (or 71 per cent) asserted confidence in their accuracy. It is likely that the question provoked an affirmative response, since it would have been difficult to acknowledge that one felt in the wrong at the moment of disagreeing. More revealing is the finding that confidence in one's accuracy increases with independence. Eighty-seven per cent of completely independent subjects asserted their confidence, in contrast with 41 per cent with 8 to 11 errors. This decrease is statistically significant ($\chi^2 = 7.7$, $p < .01$). There is equally a significant difference between those with 0 to 2 and those with 3 to 11 errors.

It is not surprising that those who were independent should be more confident in their rightness than those who went with the majority. But we also find independents who were by no means assured; the mere fact of independence did not make them immune to doubt. Correspondingly, a substantial proportion of the strongly yielding felt the group to be wrong on the few trials in which they were independent. Often they were more vehement in the interview about their independent responses than far more independent subjects. One may ask: If convinced of the wrongness of the majority, why did they follow it? This is of course one of the central problems growing out of this investigation.

The following question concerned the presence of doubt during the entire course of the experiment: "Still thinking back to the time when you were giving your judgments, would you say that the others made you doubtful about your accuracy?" Fifty-four subjects answered the question in this form. Seventy per cent acknowledged doubt. Once more we find that subjects became more doubtful with decreasing independence. ($p < .10$). Independence, however, did not guarantee protection from doubt.

In agreement with the preceding results were the replies to the following question: "Would you say that you were tempted at times to answer as the others did?" The replies of 107 subjects show that the admission of temptation rose steeply with yielding; 45 per cent of the independent (0 to 2 errors) and 85 per cent of the yielding (3 to 12 errors) subjects responding affirmatively, a difference that is highly significant. Again there were independent subjects who admitted to temptation and, more anomalous, yielding subjects who denied it. Lack of candor is the only explanation we can offer for the latter response.

One question in this series revealed no difference between the contrasted groups: "Would you say that you were concerned about the disagreements?" This question was put in an early version of the interview to 44 subjects. Eighty-two per cent acknowledged concern, and there was no relation between the replies and quantitative performance. We conclude that the fact of disagreement per se produces concern, and that concern is not a sufficient condition of yielding.

A rather interesting result was obtained in response to the question: "Did you find the judgments difficult?" Yielding subjects responded in the affirmative more frequently (significant at the 5 per cent level). One might suppose that they pleaded difficulty as an excuse for their indecisiveness. It is equally probable that the pressures under which they worked confused them to the point of actually heightening the difficulty of the task. A point worth mentioning, although it lacks statistical reliability, is that the intermediate subjects, with 3 to 7 errors, tended to find the judgments most difficult. This would be in accordance with the view that the extremely yielding subjects had made up their mind to yield and that the difficulty of judgments played a subordinate role.

Of 47 subjects who were asked "Did you doubt your own vision?," 40 per cent answered affirmatively. The same general tendency reported above was present here, too; doubts of vision increased with errors, but the differences did not reach statistical significance.

To summarize: (*a*) Most subjects felt concern over the disagreement. Concern bore no relation to performance. What differentiated between the subjects was not the presence of concern but the ways in which they dealt with it. (*b*) Conviction of rightness, freedom from doubt, and absence of temptation were all a function of independence in the experimental situation. Those with substantial errors were on the whole less confident of their rightness, more doubtful, and more tempted than the independents. However, many independents were not free of these difficulties, again suggesting that it was not the experience of conflict but the manner of coping with it that was decisive. (*c*) Yielding subjects tended to find the psychophysical judgments more difficult than independent subjects. Intermediate subjects showed a tendency to report the most difficulty. (*d*) Most subjects did not suspect that the majority judgments were not genuine. Suspicion at times occurred only as an hypothesis which, like many others, was rejected without altering the individual's direction.

---

[20] For obvious reasons this question was not put to those who erred without exception.

The Growth of Self-Reference

One important consequence of the contradiction, further aggravated by the requirement to respond publicly, was to center the subjects upon themselves, arousing a fear that they were suffering from a defect. Many concluded that their vision or judgment was misleading them, and in addition feared disclosure of their defect. Numerous statements testify to this reaction: "I felt like a silly fool" [3]; "A question of being a misfit. They agreed —the idea that they'd think I was queer" [8]. "It made me seem weak-eyed or weak-headed, like a black sheep" [3]. Some began to wonder whether the situation was not likely to turn up some unsuspected defect in them. "I felt conspicuous, going out on a limb, and subjecting myself to criticism that my perceptions, faculties were not as acute as they might be" [3]. One feared, not disapproval, "but that they would feel sorry for me" [1]. Some referred primarily to the discomfort of being the center of attention. "You have the idea that the attention of the group is focused on you" [0]. "They'd all begin to say the same thing. I'd be tempted to avoid attention and curiosity that would be aroused by consistent disagreement" [1]. I felt I wanted to go along with the crowd. I didn't want to seem different; at the same time I felt the need to give the right answer" [12]. In some cases the desire to avoid attention became compulsive. One subject described how he wanted to be independent, but "the closer it got to me the greater the compulsion [not to differ]" [3].

Some, including the most independent, feared that the group would disapprove of them for standing out or would suspect them of exhibitionism or wilful stubbornness. "I felt like Malik or Molotov" [0]. It may be of some interest that our independent subjects did not claim that they felt like Thoreau or Emerson. "The stigma of being a nonconformist—being stubborn" [0]. "Felt I was a radical—different from others" [2]. One completely independent person opened the interview with: "I hope you didn't think I was different—I was just calling them as I saw them." "I felt they'd think I was a wet blanket, or sore thumb" [3]. "I felt that I wanted to be honest and also that I must have been wrong. I like to be one of the boys, so to speak. I don't want to seem an imbecile—but then I also didn't want to seem absolutely dishonest. Was trying to see their lines as correct but succeeded only slightly, because there was always my line" [4]. "They might feel I was just trying to be out of the ordinary, objectionable, by continually being different from them. They might think it was stubborn of me" [5]. "They thought perhaps that I was responding out of spite, or that my perceptual process was not good" [4]. "Thought there was something wrong with me and wouldn't want to show it. Later I felt the group was wrong, but because I was not certain of being right I went along" [6]. "They probably think I'm crazy or something" [3].

In this connection many stressed the sense of loneliness at being separated from others. This feeling merged with an oppressive sense of the contrast between the apparently supreme security of the majority and their own bewilderment: "So many against me—so many sure of one thing. For a while it made me feel funny; it seemed as though I was a fool but I answered the way they looked" [7].[21] "They thought perhaps I couldn't see well. They probably didn't think they were wrong because they were in the majority" [7]. "It seemed simple for everybody else, complex for me" [3]. The reactions described in this section came most frequently and with a stronger charge of feeling from subjects who had erred.

---

[21] Shortly we will consider the frequent occurrence of such statements in which yielding subjects speak as if they had been completely independent.

The contradiction from the majority, which first produced a series of cognitive reactions, such as perplexity and doubt about the situation at large, eventually aroused a number of emotional reactions centering around the self. As the disagreement persisted many began to wonder whether it signified a defect in themselves. They found it painful to be (as they imagined) the focus of attention, in addition to which they feared exposure of their weakness which they suspected the group would disapprove. These circumstances fostered an oppressive sense of loneliness which increased in prominence as subjects contrasted their situation with the apparent assurance and solidity of the majority.

Stated Reasons for Independence and Yielding

The changes described in the preceding section are closely connected with the reasons subjects advance for their actions. The statements of those who were independent take a rather simple form. They either assert that they felt they were right or, in cases where they were filled with doubt, that they felt under obligation to report what they saw. In contrast, the statements of yielding subjects are more varied.

Probably the most compelling reason for yielding was the intolerableness of appearing different from the group when to do so had the meaning of exposing oneself to suspicion of defect and disapproval. "I felt the need to conform. . . . Mob psychology builds up on you. It was more pleasant to agree than to disagree" [10]. "I agreed less because they were right than because I wanted to agree with them. It takes a lot of nerve to go in opposition to them" [10]. "It is hard to be in the minority" [7].

At times errors were justified on the ground that the group was itself following the first person. "The more conformity among the group, the more it made me conform" [10]. "I think there is too much following the leader in a group like that . . . they were going along" [12]. "A lot of them just copied what the other one said. . . . I felt they weren't sure themselves and were just copying" [5]. "If people had been alone, there wouldn't be so much agreement. Some people followed" [10]. Some yielding subjects started with this assumption, later abandoning it in favor of the view that the majority was serious: "At first I thought they felt the way I did—they didn't want to be individualists. Then later it came to me they answered the way they thought was right" [12]. Independent subjects mentioned the suspicion that the group was simply following its first member somewhat less frequently, but it had a wholly different meaning for them, strengthening them in their resolve not to be weakened: "I felt some of those guys must have seen it as I did and they were being pretty spineless not to say what they saw" [0].

Occasionally one finds conventional statements about conformity that sound like faithful reproductions of "principles" of social psychology: "We all want to be with the bandwagon." "When in Rome you do as the Romans" [10]. "You always like to go along and be like everybody else" [9]. These statements, as well as the one following, indicate clearly that some subjects had lost sight of the need to report their own observations: "Some of the time I figured, 'What's the sense of my giving another answer?' If they are wrong, then I'll be wrong too, and if they're right, I'll be wrong" [8]. Finally, there were not infrequent instances of subjects who could not or would not state the effect of the majority upon them.

As the preceding discussion suggests, errors were frequently strictly conforming responses, going contrary to what the subjects clearly saw. There were, in addition, errors of a somewhat different source. Those who became convinced that their vision or judgment was awry made efforts to see the relations as the majority presumably did. Characteristic are the following comments: "I tried to make them look right [i.e., like the majority!], but they still looked wrong" [10]. "I tried to shrink the line mentally" [8]. Most subjects did not succeed, in view of the rigidity of the task, in the effort to restructure. Others, however, did apparently become confused and less well able to judge; instead of conforming contrary to their observation, these individuals became doubtful and gave the benefit of the doubt to the majority.

For this assertion we have only indirect evidence, although it seems quite convincing. As mentioned in the account of the procedure, we confronted subjects in the course of the interview (following the disclosure) with some of the comparisons they had judged wrongly. At this point judgments were uniformly correct. But at times subjects expressed strong, and to all appearances, genuine astonishment when informed of the estimates they had given fifteen to twenty minutes earlier during the experiment. One subject, looking at the stimuli of Trial 1, stated: "It doesn't look at all like 1—I don't see how I possibly could have said 1" [10]. Another, in response to the same comparison, stated: "Can't believe I changed on that" [3]. In response to the stimuli of Trial 2, another subject remarked with surprise: "The difference is very clear now" and although he had erred three times, he continued to insist that "I went with the crowd once—only once in the middle." Still more impressive were the reactions of a very few subjects who stoutly, and with great seriousness, maintained to the end that all the estimates they reported had seemed to them clearly accurate at the time. One subject with six errors stated: "I always picked the line I thought was equal. Some seemed questionable, but the one I chose seemed to be the closest one to it [the standard]. I always thought I was right, but I wasn't sure." In the same vein, a subject with three errors stated: "Toward the end I began to doubt myself but even then I stuck to my guns." Knowledge of their past estimates did not alter the opinions of these subjects. It was the writer's impression that they were reporting what they believed to be a fact. We are in the dark as to how the field of these individuals might have become blurred to this point. (See also pp. 42–43 for further discussion.)

Awareness of Independence and Yielding

In general, the independent subjects were more frank and realistic. They frequently showed a clear awareness of what they had been doing. This was especially evident in the cognizance they took of their occasional errors, which they acknowledged and discussed in a forthright way, in contrast to yielding subjects who so often minimized them and attributed them to circumstances outside themselves.

For example, one subject remarked concerning his only error: "I had not decided [upon the answer] before they started. Immediately after [responding] I thought that I had been wrong—but it was too late."[22] The following was a reaction at times encountered in younger subjects with intermediate errors: "After I did it [yielded] I was sorry; it seemed to be a stupid thing to do. If you see it one way you might as well say it" [3]. "A couple of times I knew that I was weakening, and was peeved at myself" [3]. Another subject remarked that he noticed his errors and "demanded of myself to have the courage of my convictions. I pulled myself together and said: 'Regardless of what the others say I'll report what my eyes tell me.' I had to build up a certain amount of defiance to give a different answer" [5]. Similarly, at the point of disclosure, the more independent subjects were not above giving way to their pleasure and shocked surprise. "You have doubts when you finish a thing like that," remarked one subject gratefully. "I am elated and surprised" [3]. Even when they were on guard, their reactions had a quality of candor, as did the following statement by a subject who apparently feared to expose his feelings to further shock: "I'd want more proof, definite proof that what you were telling me was right" [1].

The reactions of yielding subjects were more often evasive and shallow, and some revealed until the end a lack of appreciation of the situation and of the possible significance of their action.

When asked to describe his experiences at the outset of the interview, one subject with twelve errors inquired: "Exactly what do you mean by experiences?" Another remarked: "I didn't have any experiences—I felt normal" [8]. All that another could find to say at this point was: "It was frustrating—I still think I was right" [10].

---

[22] The present conditions permitted the subject to inspect the lines and match them before the majority started to announce its estimates. As the experiment progressed, and as they became aware of the potential power of the majority, many subjects took advantage of this opportunity. It was, for example, quite usual for them to mention that they were determined to reach a judgment before the group and to adhere to it. "I was confident at first, then became doubtful. Then I made my decision before I heard anyone else and stuck to it to remove the danger of following the leader." "I judged before the first man and stuck to it." It is our impression that this opportunity for independent prior judgment had a considerable effect in heightening independence. Unfortunately we have no means of telling whether or how subjects differed in readiness to take advantage of this possibility. Relatively simple experimental variations could clarify this question. The procedure could be modified in such a way that each critical subject would hear the majority opinion before reaching his own conclusion; also one could lengthen the interval between exposure of the cards and the time when the majority begins to respond so as to insure that each subject compared the stimuli before hearing the discordant judgments. It might furthermore be of value to vary systematically the length of time during which the subject remains alone with his own judgment.

Confronted with his earlier estimates, a completely yielding subject remarked: "If I knew we were supposed to disagree, I would have. I thought something was wrong with their eyes." The only explanation for the disagreement that another ventured in the course of the interview was: "People perceive things differently" [10], and still another, "How do we know who is right" [10]?

Yielding subjects frequently denied that the majority had exerted any effect upon them. "Because everybody else was answering in a certain way, I wasn't going to answer the same way. I felt I was right and that's all there was to it" [12]. This subject was referring to his "independence" on the neutral trials, slurring over his conduct on the critical trials. "I said what I saw; didn't want to invalidate the experiment" [9]. One subject, with nine errors, summed up his reaction at this point thus: "I feel rather pleased; I'm interested in psychology." We often observed a similar lack of opennesses to the situation in the reactions to the disclosure. "I always thought my judgment was right—I felt peculiar, but I knew I was right" [6]. To the first step of the disclosure (see p. 25) another subject who erred 11 times remarked: "I'd be proud of myself." Only with the final disclosure did he say: "Well, I'll be quite frank with you. Naturally I didn't know what the story was—and I did follow the leader too much. I could have been more honest with myself, but there were six ahead of me. I guess there were a few in which I could have answered the other way." Another subject was still troubled, after the disclosure, with the thought that "perhaps I was doing something wrong by not going along with the crowd" [6].

Only with the final disclosure did yielding subjects speak more openly. Most remarkable were the comments of gratification with which some greeted the information. They were prone to seize upon the few instances of resistance as a demonstration of their steadfastness and to banish the others from awareness. One subject, with seven errors, remarked: "I would be happy" [Why?] "It's not very often you find one right and eight wrong. It's good to find out you're right—especially when so many disagree with you." "So I wasn't seeing double!" [6]. I'm gratified to a certain extent that my decisions weren't incorrect all the time" [8]. Understandably some could only feel relief at finding that their private judgments had been correct. There were also comments such as the following, which revealed an effort to draw a moral: "I was affected both by doubt as to what was the right answer and by group pressure. This will teach me a lesson to stand up for what I think is right. You can guess now yourself that it was mostly group pressure that influenced me—I'm a little ashamed of myself" [12].

Yielding subjects underestimated their errors to a remarkable degree. It was not unusual for them to report that they erred once or twice when they had responded with the majority ten or twelve times. "I think I followed a couple of times . . . on a couple of occasions I decided to be a conformist" [10]. "Two times, to be precise" [10]. "I was doubtful . . . I went with the group about twice; I hadn't made up my mind—it was a toss of a coin" [10].

Concerning this reaction we also have some quantitative data. Fifty-eight of the subjects answered the following question: "Did you ever answer as the others did against your own first choice?" If the answer was in the affirmative, it was followed with the question: "How often do you remember doing so?" Table 12 is a scattergram plotting the estimated against the actual number of errors. Correct estimates should cluster around the diagonal line. What we find is that, except for the independents who were almost completely accurate, only two other subjects gave a correct estimate. All others underestimated, and the underestimation increased with the actual number of errors. The mean discrepancy between actual and estimated errors was 4.0 ($t=7.14$, $p<.001$).

What might be the reason for this extreme underestimation? We can rule out forgetting in the usual sense; if it were a question of retention, we should also have found overestimations. To make doubly sure whether a general function of memory was involved, we compared the preceding results with those obtained in response to the following question that was put to another portion of the population, 49 in number: "How many times did you disagree with the group?" The results appear in the scattergram of Table 13. Here we find again a trend to underestimation, but far weaker than in the preceding case, the mean discrepancy between actual, and estimated disagreements being .92. More of the estimates were now correct, and there were a few overestimations. The discrepancy between actual and estimated disagreements is also significant ($t=3.40, p<.001$). But the tendency to underestimate yielding reactions was significantly higher than the underestimation of disagreements ($t=4.41, p<.001$).

The data, while establishing the fact of underestimation, do not enlighten us about its source. On the basis of general observation we are inclined to believe that the bulk of underestimation resulted from the subjects' unwillingness to admit this unpleasant fact. Had we repeated this question following the disclosure, it is likely that the underestimation would have been lowered. It seems to us that this also accounts for the trend to underestimation of disagreements. Some subjects might have tried to avoid the appearance that they had been too much at odds with the majority. This was not, however, the sole factor at work. Undoubtedly some individuals were confused to the point where they became far less certain than they normally were.

TABLE 12

SUBJECTIVE ESTIMATES OF ERRORS*

($N=58$)

Estimated number of errors	0	1	2	3	4	5	6	7	8	9	10	11	12
12													
11													
10													
9													
8													
7													
6										1			
5													
4										1			
3				1		1	1		2	1			
2				4		1	1	1	2			1	
1	1		5		1	2	1						1
0	17	1	2	2	1	1	1	1	1	1			1
Actual number of errors	0	1	2	3	4	5	6	7	8	9	10	11	12

* For the purpose of this analysis compromise responses were categorized as disagreements.

What can be said concerning the reality value of the situation? For most it was a situation of appreciable tension, although there were important differences in responsiveness to the majority. Some felt the conflict more keenly than others; divergences in this respect may have been significantly connected with independence and yielding.

The evidence cited above indicates that generally the subjects acted as if something of consequence were at stake and that neither independence nor compliance were trifling matters. Concern, doubt and temptation were the prevailing reactions and were expressed in terms of fairly strong emotion. The fact seems to be that the circumstances quickly gained an im-

TABLE 13

SUBJECTIVE ESTIMATES OF DISAGREEMENTS*

($N=49$)

Estimated number of disagreements	0	1	2	3	4	5	6	7	8	9	10	11	12
17													1
16													
15											1		3
14													1
13													
12											1		3
11													1
10											1	1	
9									1		1	2	1
8													1
7								1	1	1		2	
6							1	2		1	1		2
5									2	1			1
4						1		1		1			1
3					2								
2			2	3	1								
1		1	1										
0													
Actual number of disagreements	0	1	2	3	4	5	6	7	8	9	10	11	12

* For the purpose of this analysis compromise responses were categorized as disagreements.

port that went beyond the immediate task. The subjects had undertaken to report correctly on simple facts, a requirement not easily confused with observing the social amenities. In confirmation we may observe that our subjects almost universally evaluated independence positively, regardless of their performance. None sang the praises of conformity, and even when they did not live up to their precepts, they did not question that there are criteria of right independent of, or contrary to, group standards. Consequently, those who were independent expressed relief and joy when informed of the circumstances, while those who were caught by the majority were rueful but never proud. In fact, the situation possessed personal relevance because it provoked a conflict of values and revealed to some a double system of values at work—one value which prompted them to conform, coupled with another that asserted the importance of thinking for oneself and being an individual.

Less responsive subjects spoke as follows: "I didn't go along because I had nothing to lose by sticking to my own perception" [1]; "It was a psychological experiment, not like a life-or-death situation. I thought there was some importance attached to it—and that I'd take a scientific attitude toward it" [0]. A few yielding subjects spoke in the same vein: "I'd see one, then they'd answer, and I'd figure it wasn't important enough to stand out; I thought, 'Oh, what the heck, I'll just agree with them' " [10]. Or, "This seemed of no importance, and I'd stand out. It had no real bearing and didn't make any difference" [9]. These comments were not representative, and in some cases they did not describe adequately the situation of the subjects from whom they came. For example, the subject responsible for the last statement was affected to the point of stuttering during the interview.

C. Forms of Independence and Yielding

The great differences in independence among persons require us to take the individual as the unit of observation and to describe the major forms of reaction to the experimental conditions, in contrast to the procedure, followed above, of attempting a general sketch that would apply to most subjects. Accordingly, we will now describe a number of distinguishable modes of reaction illustrated by a few selected subjects. For this purpose we have chosen subjects from the extremes of independence and yielding; as far as we could observe, intermediate subjects did not reveal notable characteristics absent in the extremes.

It might appear on first thought that the grounds of independence would be substantially similar, as would the grounds of yielding, and that subgroups selected from the extremes would be internally homogeneous and show uniform contrasts when compared. As the experiment progressed, however, it became evident that the grounds of independence, as of yielding, were diverse, and that individuals with similar or identical performances (in terms of errors) often differed strikingly in their psychological reactions. At the same time the modes of reaction did not vary indefinitely; certain of them repeated themselves. We have accordingly singled out for description a number of outstanding types of reaction in independent and yielding subjects. The basis of our categorization is provisional and must be taken as a first approximation.

Forms of Independence

The independence of strength. The outstanding characteristic of subjects in this group is their capacity to retain faith in their experience in the face of massed opposition. At the extreme end we find that the sense of confidence forestalls the growth of conflict and places the majority in a dubious light. So much is this the case with a few that they even appear to enjoy their situation somewhat. This is by no means, however, the usual reaction of independent subjects. Most frequently they are responsive to the majority and feel keenly their strange position. This is evident in their frank acknowledgment of doubt and temptation and in the outspoken gladness with which they greet the explanation of the experiment. What is most characteristic of the present group is their capacity to withstand without substantial difficulty the doubts and the loneliness of their situation.

Subject 39 was completely independent. He believed the majority to be wrong, possibly subject to an illusion of which he was free, or victim to the suggestion of the first member. He was unusual in finding the situation somewhat pleasant, deriving some amusement from maintaining an independent position. But his enjoyment was clouded by occasional concern that *he* might be the one subject to illusion. This doubt did not disturb him much, however, since he felt that the probabilities were strongly in his favor. He was therefore prone to look upon the majority as "spineless" followers who contradicted their convictions. It was our impression that this subject's assurance freed him to reject the group pressure and forestalled the growth of conflict. The following is an excerpt from his interview.

Did the fact that other people were present and attending to the same task affect you in any way? *No.*

How many times did you disagree with the group? *Fifteen times.* (Unlike most, he overestimated the frequency of disagreement.)

When you gave an estimate that disagreed with the others, did you feel that your perception was objectively correct? *Objectively correct.*

Did you feel that the perceptions of the others were objectively correct? *They were wrong. Or, if I was wrong, I'd rather try to find out why I was wrong!*

How did you feel when you continued to give answers different from the others? *Mixed emotions—would be amusing if I turned out to be right, but if wrong, sort of lonely feeling.*

What did you think when you first heard the others give an answer different from the one you were going to give? *My first impulse was that maybe they didn't see that one well.*

Did you examine the lines more closely? *Yes.*

Did you feel any differently about the later disagreements? *Yeah, I figured the lines contained some sort of illusion which I was not subject to and they were.* (This statement carried the implication that his judgments were superior.)

Did the others arouse doubt concerning the correctness of your judgments at the time you were giving your estimates? *Yes.* (Adds that the doubt was moderate.)

What was the nature of the doubt? *Maybe I was subject to illusion.*

Did you have doubts that your eyes might be deceiving you? *Two only—a couple of times, then I was sure I was right.*

Were you concerned about the disagreements? *No, the whole thing was not too important. I'm used to being different at times.*

How did you feel when the others and you agreed—when they gave the same answer that you had in mind? *That's what made me think it was some kind of an illusion. They seemed to get the short ones right.*

Did you feel that the answers of the others were arousing a tendency in you to agree with them? *No.*

As you continued to look at a pair of cards, did the answers of the others continue to appear wrong or did they seem to become more plausible? *I actually thought the others were subject to the suggestion of the first one.*

As you continued to hear the answers of the others, did the discrepancy between their answer and the one you were going to choose remain the same, or did it change? *No, I was getting almost a sadistic pleasure out of being different.*

What did you think about the third line, the one no one matched with the standard? *Nothing.*

Did you think the group would disapprove of you, or think you were peculiar if you gave a different answer? *Not disapprove, but they have a habit of laughing at you if you're wrong in class, but in this case I didn't care. It would be different if it were a question of ethics, but I wouldn't agree!*

What did you think our purpose was in this experiment? *Just vision.*

At the beginning, after the first three or four judgments, were you aware that you were in disagreement with all the students? *Yes.*

Were you aware that all the students were in agreement with each other? *Yes.*

Did the disagreements produce in you the feeling that you were separated from the group, or in opposition to the group, or excluded from the group? *Psychologically, yes.*

Did it require an effort to answer differently from the others? *No.*

When you answered differently from the others did you feel that you were resisting them? *Slightly, only.*

Did you expect, as the experiment progressed, that they would continue to be in agreement with each other? *Yeah, after two-thirds through. Also thought they might break up.*

Did you feel an increasing sense of opposition as your turn approached to give a judgment? *No.*

Suppose I told you that your judgments were always right when you disagreed with the group, what would you think? *Not too much surprised. I'm used to having what I see be correct.*

The subject's reaction to the full disclosure was mild, although he was happy and interested. Subsequently he reported no suspicion of the majority. We now cite a few of the comments following the disclosure.

Were you self-conscious at the beginning of the experiment? *No, more towards the end. I figured somebody else would be disagreeing soon.*

Did the disagreement of the group produce in you a feeling of irritation? *Yes, slightly. I felt some of those guys must have seen it as I did and they were being pretty spineless not to say what they saw.*

What in your character and experience would you say was responsible for the way you acted in this experiment? *Long years of practice in being different from other children. I've never had any feeling that there was any virtue in being like others. I'm used to being different. I often came out well by being different. I don't like easy group opinions.*

In the following completely independent subject we have a person equally independent but decidedly more affected by the majority. While he was strongly task-oriented he also entertained some doubt corresponding to the actual uncertainties of the situation. The doubt provoked a genuine emotional reaction which, however, did not deter him from independence.

Did the fact that other people were present and attending to the same task affect you in any way? *I thought there must be something wrong—and I thought for a second of going with the group.*

Would it have made any difference if you had been here alone? *No; except for a few seconds.*

How many times did you disagree with the group? *About 60 per cent* (of the trials).

When you gave an estimate that disagreed with the others, did you feel that your perception was objectively correct or incorrect? *Yes—completely obvious.*

Did you feel that the perceptions of the others were objectively correct? *Well, let's see—it would naturally follow. But I didn't feel that their perceptions were incorrect, rather that there was something wrong.*

How did you feel when you continued to give answers different from the others? *Something was wrong with either my perception or with all the others.*

What did you think when you first heard the others give an answer different from the one you were going to give? *That they were wrong, that the lines were throwing off their perception.*

Did you examine the lines more closely? *I first made my judgment; then I studied the lines.*

Did you feel any differently about the later disagreements? *The thought struck me that they were all wrong and I was the only one right—but that was a momentary thought.*

Were you concerned about the disagreements? *No, except to wonder why there was a disagreement.*

Did you feel that the answers of the others were arousing a tendency in you to agree with them? *No.*

Were you tempted at times to answer as the others did? *Only once.*

Did the fact that the others agreed with each other make them appear as a group? *Yes—only insofar as their perception of length was the same.*

Did they as a result appear closer to each other than to you? *Yes—there was some of that feeling—sort of being on the edge.*

Did they seem to become more united as the experiment progressed? *Yes.*

Did the disagreements produce in you the feeling that you were: (*a*) separated from the group? *Yes.* (*b*) in opposition to the group? *Yes—insofar as the judgments were concerned.* (*c*) excluded from the group? *No.*

Did it require an effort to answer differently from the others? *No.*

When you answered differently from the others did you feel that you were resisting them? *No—I was reporting what I saw.*

Suppose I told you that your judgments were always right when you disagreed with the group, what would you think? *I'd want more proof, definite proof that what you were telling me was right.*

Did you suspect that the group intended to affect your judgments? *Yes—once only about one fellow—not about the group.*

Did you retain this suspicion? *No.*

Did you notice whether any of the comparisons were repeated? *No.*

What in your character and experience would you say was responsible for the way you acted in this experiment? *It reminded me of a time when, as a child, my mother threw a cupful of water at me because I couldn't stop crying. My sister was there, and everyone seemed against me; I was separated. . . . I always disagreed on religion—couldn't quite swallow what was taught. . . . Have a tendency to stick to my point.*

In the reactions of the next subject, who with one exception was independent, we observe a more pronounced effect of doubt, pressure, and temptation created by the majority.

When you gave an estimate that disagreed with the others, did you feel that your perception was objectively correct or incorrect? *Yes.*

Did you feel that the perceptions of the others were objectively correct? *Yes* [with some hesitation].

How did you feel when you continued to give answers different from the others? *I'd bet on their side, because my own judgment has no better chance of being correct than theirs.*

What did you think when you first heard the others give an answer different from the one you were going to give? *I was very surprised; naturally I wanted to answer as they, but I wanted to answer accurately.*

Did the others arouse doubt concerning the correctness of your judgments at the time you were giving your estimates? *After several trials.* [He describes the doubt as moderate.] *I began to question whether my own perception was as acute as it seemed to be.*

Would you say that you were concerned? *Yes* [moderately] *because I could see how such inaccuracy could affect various fields of activity, such as driving.*

How did you feel when the others and you agreed—when they gave the same answer that you had in mind? *It bolstered my judgment—it meant that I was not completely off the track.*

Did you feel that the answers of the others were arousing a tendency in you to agree with them? *Yes, but I completely resisted it.*

Would you say that this tendency acted as a pressure on you? *I have the natural tendency not to want to be the one wrong out of the entire group.*

Were you tempted at times to answer as the others did? *Yes, but I didn't. Frankly, I considered momentarily the policy of agreeing with them and reserving my judgment—to satisfy my natural tendency not to seem different, to agree—but I decided against it.*

Did you think the group would disapprove of you, or think you were peculiar if you gave a different answer? *The test went too fast for me to develop such clear feeling—perhaps only rudimentary. Being the only one was what mattered.*

Did it require an effort to answer differently from the others? *Yes—but once my turn came, no real effort was needed.*

When you answered differently from the others did you feel that you were resisting them? *It gave me a certain sense of pride to answer differently.*

Suppose I told you that your judgments were always right when you disagreed with the group, what would you think? *I would be happy—but I would doubt you very much. The evidence is in favor of the others.* The full disclosure took him completely by surprise. When he had recovered he commented: *I am very, very glad. It relieves me of any worry, and it makes me very glad also that I gave the answers I did.* Then, after a pause: *If it was an issue more emotionally charged or more important, involving a more emotional result, I'd be much less likely to accept their judgments.* In response to the final question about the probable relation of his conduct to his character, he stated: *A certain pride I have always taken in my judgment when I was reasonably sure of it. I've always been glad to defend my opinion, enthusiastically, even when in a minority.*

The following subject, who went with the majority once, entered the experimental room in a boisterous mood, but soon became much sobered. He felt little pressure to conform, but was mildly concerned over what his judgments might do to the experimenter's results. At the opening of the interview he remarked that he felt as if in a snake pit. Quite noteworthy is the fact that he felt strong guilt over his one dereliction to which he constantly returned (in contrast to others who took continuous yielding far more lightly) with obvious feelings of self-devaluation. The entire situation seemed to be a revelation to him of the unsuspected susceptibility he harbored to group compliance. A portion of the interview follows:

Did the fact that other people were present and attending to the same task affect you in any way? *Only on one decision, which I changed, and which I shouldn't have changed.*

Would it have made any difference if you had been here alone? *Only in that one instance.*

Did you feel that the perceptions of the others were objectively correct? *It ran through the back of my mind—maybe it was lack of vision, then thought maybe I was just wrong.*

What did you think when you first heard the others give an answer different from the one you were going to give? *That perhaps other different answers would come up.* He added that he suspected, vaguely, that perhaps others were trying to influence his judgment—*the whole thing might be rigged.*

Did you feel any differently about the later disagreements? *No—not much feeling, I was just surprised and a little worried.*

Did you feel that the answers of the others were arousing a tendency in you to agree with them? *No. If it had been a more serious matter, and if the group had appeared to resent my opinion, I might have been more affected.*

Were you tempted at times to answer as the others did? *I did so, once, and was surprised that I had* *done so. I knew I was changing from what looked correct.*

Try to describe the reasons as carefully as you can. *It was the only time I considered the one they picked. Thought it might fit (for a moment), and mainly wanted to conform with the group.*

As you continued to hear the answers of the others, did the discrepancy between their answer and the one you were going to choose remain the same, or did it change? *It was still clearly different, except the one case.*

Why did you follow the group in that case? Try to describe exactly. *I wanted to conform. Was picturing in my mind the graph of results with a big dip in it—I wanted to make your results better.*

How did you feel when you saw the lines one way and answered in another way? *I was disgusted with myself for changing and almost asked for another chance at it.*

Did you think the group would disapprove of you, or would think you were peculiar if you gave a different answer? *I looked to see how the group reacted, and they didn't react strongly, didn't affect me much.*

Now we want to ask you about the time you answered as the majority did, against your first choice. As you continued to look at the pair of cards, did the answers of the others continue to appear wrong, or did they seem to become more plausible? *I made my own judgment, then considered the group's judgment, gave it, and regretted it. I never felt I was right and the group wrong, just knew what I saw.*

When you gave such an answer, did you at times forget about the question of accuracy, and consider rather that it might make you appear different from the others? *It was a question of conformity—the group didn't make me feel bad, I just wanted to agree.*

What did you think our purpose was in this experiment? *I only knew what I'd been told. I was trying to give you honest answers.*

Did this situation acquire any personal importance for you? *In what way? No, but I was afraid it might in terms of my future relationships with other students—if I were found to see poorly.*

Did it require an effort to answer differently from the others? *My answers did become less positive in tone.*

Did you feel an increasing sense of opposition as your turn approached to give a judgment? *I began to hope, towards the end, that someone would pick the one I had in mind.*

Suppose I told you that your judgments were always right when you disagreed with the group, what would you think? *I guess my answers tried to minimize my disagreement with them.*

Following the full disclosure he stated: *If I only hadn't chickened out, it would have been an even dozen.* His final comment was: *If I hadn't been required to answer, I wouldn't have done it* [yielded] *after hearing the group. It is not pleasant to be the only one different.* He then continued: *I think my father would be very sore at me for that one I missed. He would expect me to always do what I know or think is right because it is right. My mother too. Don't know why I fooled myself into thinking there might be some doubt about it.*

Independence without confidence. The subjects in this group were assailed by doubt, were convinced that their judgments were inaccurate and those of the majority right—and yet they remained independent. In terms of their feelings and ways of seeing the situation, they resembled a group of yielding subjects to be described below (see pp. 43–45). But it did not occur to them to compromise with the majority because they did not lose sight of the obligation to respond in accordance with their experience. The reason for their independence, according to their statements, was simple and uniform. They felt that they were required to respond accurately, that "if I were to be honest I'd have to say what I saw." Because they did not lose sight of their obligation it hardly occurred to them to compromise; when this alternative presented itself, they were able to dismiss it without much difficulty. Some spoke explicitly of the necessity to act independently of pressure.

This subject, aged 18, was independent on all trials but one. He was strongly affected by doubt, believing the judgments of the group to be accurate. Although he was concerned and tempted to join the majority, he yielded only once. There was only one reason he could assign for his independence—that "the experiment required an accurate answer."

Can you explain the disagreement? *I can't—unless it's a fault in my eyesight or judgment.*

How confident are you of your judgments? *In the light of the opinions of the others I'd say I was wrong, but I answered as I saw.*

When you gave an estimate that disagreed with the others, did you feel that your perception was objectively correct or incorrect? *I don't know.*

Did you feel that the perceptions of the others were objectively correct? *Probably the others were right.*

What did you think when you first heard the others give an answer different from the one you were going to give? *I first doubted, wondered. However, I felt that rather than go along I'd make the answers that appeared right to me.*

Did the others arouse doubt concerning the correctness of your judgments at the time you were giving your estimates? *Yes, sir* [considerable doubt]. *There was such an overwhelming weight of opposition against me that I wondered whether my eyesight or judgment were at fault—or an optical illusion which either they or I did not perceive.*

How did you feel when the others and you agreed—when they gave the same answer that you had in mind? *I felt that was very nice. I don't believe it was wishing to be with the herd, but I was glad there was agreement.*

Did you feel that the answers of the others were arousing a tendency in you to agree with them? *Yes.*

Would you say that this tendency acted as a pressure on you? *Yes. I had a tendency to feel that perhaps I was wrong and might just as well agree with them. As disagreement continued it looked as if I was differing either to show off, to be an individual, or trying to stand out. I did not like that.*

Were you tempted at times to answer as the others did? *Yes. But the experiment required an accurate answer.*

Suppose I told you that your judgments were always right when you disagreed with the group, what would you think? *I'd think that you were telling me that to test my psychological reaction. I felt I was right in some cases, but I'd be very suspicious if you would tell me I was right in every case.*

The following subject, age 21, yielded once; he presents a picture very similar to the one described above.

What did you think when you first heard the others give an answer different from the one you were going to give? *First I thought the others were wrong, but if I were to be honest I'd have to say what I saw.*

Did the others arouse doubt concerning the correctness of your judgments at the time you were giving your estimates? *Yes* [great doubt].

Did you have doubts that your eyes might be deceiving you? *Yes,* [or the angle].

Did this doubt change as the experiment progressed? *I became more doubtful.*

Where you concerned about the disagreements? *Oh, yes* [moderately].

Were you tempted at times to answer as the others did? *Yes.*

Suppose I told you that your judgments were always right when you disagreed with the group, what would you think? *I'd believe it.* Why? *Were the other guys suggested by the first?*

All he could say following the full disclosure was: "*I was supposed to give the right answer.*"

Forms of Yielding

The grounds of yielding to the majority were also diverse. Many subjects could not be said to belong clearly to one group or another, mainly because numerous, at times contradictory, reasons and motives played a part in their reactions. Before describing some of the predominant reactions, we present below the case of one subject who illustrates the varied phases of the problem.

This subject went with the majority without exception. He was deeply affected but tried to describe his reactions to the best of his ability. Because he went completely with the majority, some of the prepared questions were not suitable; we therefore questioned him in a more individual manner.

He spoke at first with difficulty, opening the interview with the statement: *If I'd been the first* [to respond] *I probably would have answered differently.* This was his way of stating that he had knowingly adopted the majority estimates. But he underestimated the frequency of his errors. "How often did you answer as the others did, against your own first choice?" *Possibly as many as one-fourth or one-third. Mostly I wasn't sure, I was undecided. If so many people say one thing it is bound to influence you.* He claimed, as many others who erred, that some of the comparisons were difficult. Asked why he answered as the group, he replied: *Pure habit—sheer force of habit. On the doubtful cases I would go along. I felt they were probably wrong, but I wasn't ever absolutely sure. I didn't think they were right—it was mostly pretty much of a toss-up. I didn't have time to think it over, and they answered so fast; if they had been doubtful, I probably would have changed, but they seemed to be absolutely sure, and you tend to follow.* In further explanation of the acquiescent response he added: *Some kind of pressure builds up in you. On the first one I almost started to say something different but afterward I more or less fell in with them—you know, "hesitate and all is lost.'"*

Following the disclosure he volunteered: *I suspected about the middle—but tried to push it out of my mind.* The information interested him and he became eager to help with his observations, adding spontaneously: *If I had known more about it, if I had a personal interest in it, maybe it would have made a difference—I didn't know what it was for.*

Did you think the task was not important? *I didn't consider the importance, not terribly important . . . I just sort of slipped along.*

Did you consider that we wanted your own judgment? [embarrassed]: *I didn't think of that; if it entered my mind I more or less pushed it out.* Did you suppose that they would disapprove if you disagreed? *I don't think so. It might have been subconscious, but I didn't consciously think they'd think I was queer; it might be . . . but I wouldn't be afraid to. In a thing like this they would be more likely to be right; it's cut and dried. I wasn't giving a terrific amount of thought to it—I don't think the majority is always right.*

He was now confronted with some of the stimulus lines and informed of his earlier estimates. In response to Trial 1 he stated: *I tried to make myself see it as equal, and succeeded to a more or less degree. I thought maybe the numbers* [beneath the lines] *or the arrangements of the lines had something to do with it. At first I also thought maybe it was an illusion, but abandoned that idea. I thought they might have been seeing an optical illusion that I didn't. Then I thought that the ones ahead of me were following the ones ahead of them . . . just like I was. There was no time to draw a conclusion. It wasn't that I actually saw it this way—I just fell in. On the doubtful cards the fact that they saw it before made it doubtful. If they hadn't been there, if I had been here by myself I would have been pretty positive.* He concluded with the following remarks: *I didn't see how they could have answered so quickly when I was so doubtful, though I probably would have agreed anyway. If it had been a political question, I don't think I would have agreed if I had a different feeling. I probably wanted my own ideas, but it was easiest to string along. Even though I knew that pressure was there, it's funny that I would give in just like anybody else.*

This protocol is of interest for a number of reasons. It illustrates the many-sided considerations that played a part in the reaction to the episode. In one respect it is not typical; this subject did not report much concern about the group's opinion of him. Further, the reasons for his action do not become fully clear. The impression one gets is that he was caught off guard, was overwhelmed by the apparent assurance of the majority, and allowed himself to become uncertain about many of the judgments. Particularly noteworthy is that he continued with the majority although he was not convinced it was right, thinking that it might be subject to an illusion of which he was free, and suspecting the group of following out of expediency. At the same time one gets a strong sense of the presence in this subject of forces contrary to conformity.

Yielding at the "perceptual" level. Among the rarest, but not the less significant, of the reactions were those in which yielding occurred apparently without awareness. Previously we have described the case of one such subject who yielded on all trials but one, but who asserted with every appearance of genuineness that he never reported an answer he believed to be wrong (1, p. 469). The most noteworthy feature was the placidity of the subject; it appeared that his lack of awareness enabled him to maintain an attitude of complacency and frankness, only mildly troubled by the suspicion that something unknown to him had occurred.

The writer has not himself observed other cases as extreme as this, although a few were reported by his associates. In these instances there appeared to be a genuine lack of awareness of the degree of yielding, evidenced by an insistent rejection of its possibility. This reaction was present, however, together with others, such as a deep desire to join the group and a sense of personal inadequacy. For lack of a direct test we cannot conclusively establish the fact of unawareness, although the subjects did convince us that they were trying to report truthfully. Also, in the nature of the case, we cannot establish, assuming the absence of awareness, the level of distortion. In the absence of evidence it would be too far-fetched to suppose, mainly because we have no means of comprehending, that there was an actual distortion in the perceived relation of lengths. It is more likely that the distortion occurred principally at the level of judgment, without the subject's knowledge of his contribution. To put the matter somewhat differently, these subjects granted to the majority the power to see correctly, attempted to see the relations as the majority did, allowed themselves to become confused, and at the critical point adopted the majority judgments without permitting themselves to know of their activity.

In a more moderate degree there was cognitive confusion in a larger number of subjects, as we have mentioned earlier (see pp. 32–33). We are referring to instances in which subjects, upon confrontation during the interview with the comparisons and their earlier pro-majority responses, found it difficult to understand how they could have deviated to the degree they did. Again, for lack of direct tests, the evidence is not conclusive. It is possible that the subjects yielded in a more conscious way but veiled the fact from themselves after the lapse of 15 to 20 minutes. Assuming that they were trying to report correctly, it would be a necessary inference that inhibition of awareness occurred at the time of recall.

Yielding at the level of judgment. The subjects included in this group were robbed of confidence by the opposition and quickly reached the conclusion "I am wrong, they are right." They spoke much like the independent subjects without confidence (see pp. 40–41), but unlike the latter they lacked any other source of support. To be sure, they did not like to expose their judgment, but this was not the sole or even the main reason for yielding. Rather their mental field was narrowed down to a preoccupation with their accuracy (or lack of it) to the point where they appeared to feel that they had lost the right to express their inferior judgments. This is the only way in which we can understand the recurring statements that they went with the majority because they did not wish to spoil the experiment. These reactions suggest that the presence or absence of confidence alone are not decisive, that equally important are the resources for coping with unavoidable doubt and strain.

The subject, age 20, erred ten times. His principal concern was to report accurate judgments, and he adopted the judgments of the majority solely because he was convinced they were right and superior to his. Because of poor coordination of his bodily movements he had come to expect that he will usually be deficient in tests of physical effectiveness. He suffered no cognitive confusion, continuing to see the relations clearly in his own way. He also claimed that he did not mind differing from the group.

Did the fact that other people were present and attending to the same task affect you in any way? *Only insofar as when I was doubtful I thought they were correct. It didn't bother me to disagree with them, or I wouldn't have done it at all.*

When you gave an estimate that disagreed with the others, did you feel that your perception was objectively correct or incorrect? *I gave the answer I saw.*

Did you feel that the perceptions of the others were objectively correct? *Yes.*

What did you think when you first heard the others give an answer different from the one you were going to give? *I just wondered what was the matter with me.*

Did the others arouse doubt concerning the correctness of your judgments at the time you were giving your estimates? *Yes.* [Moderate doubt.]

How did you feel when the others and you agreed—when they gave the same answer that you had in mind? *Until I'd given a different answer, it didn't matter, but after that, it was reassuring.*

Did you feel that the answers of the others were arousing a tendency in you to agree with them? *Only in four or so cases. Only when really doubtful, then I'd agree. I thought they were probably right, nothing more.*

Were there times when you saw the lines one way and you answered in another way? *In those four to six cases I agreed because I figured they were right.*

Did you ever give an answer which you knew to be wrong? *No. I only assumed my answers to be wrong, because I disagreed with everyone else.*

Did you think the group would disapprove of you, or think you were peculiar if you gave a different answer? *No—if I had, I wouldn't have differed on any of them.*

What did you do about your first choice, the one you thought was really correct? *I tried to compare mine—I gave the group answer against the card. Often mine still looked best, but I figured they were right.*

When you gave such an answer, did you at times forget about the question of accuracy, and consider rather that it might make you appear different from the others? *No, it was always a question of accuracy.*

Suppose I told you that your judgments were always right when you disagreed with the group, what would you think? *Hm-m-m. I wondered if it was a put-up job, because those guys were all up here when I arrived.*

The next subject, age 19, yielded ten times. He quickly became tense and worried, examining the lines more and more closely. The fifth comparison, which was neutral, obviously relieved him. Then he went along with the majority in a high-strung, shaken way.

Did you feel that the perceptions of the others were objectively correct? *Yes.*

What did you think when you first heard the others give an answer different from the one you were going to give? *That they were wrong.*

Did the others arouse doubt concerning the correctness of your judgment at the time you were giving your estimates? *Yes.* [Moderate doubt.] *The majority was against me—I thought perhaps I was wrong.*

How did you feel when the others and you agreed—when they gave the same answer that you had in mind? *I thought we must be right.*

Did you feel that the answers of the others were arousing a tendency in you to agree with them? *Yes.* [Moderate doubt.]

Were you tempted at times to answer as the others did? *In one case there was a slight bit of doubt. It was rather close and I wasn't sure of myself.*

As you continued to look at a pair of cards, did the answers of the others continue to appear wrong, or did they seem to become more plausible? *Except for the two occasions the group answers seemed plausible.*

So convinced was this subject of the accuracy of the majority that he could not respond adequately to the first step of the disclosure. "Suppose I told you that your judgments were always right when you disagreed with the group, what would you think?" *I'd think something was fishy.* He received the full disclosure in a controlled, inhibited way, appearing guilty and unable to admit the fact of yielding. Toward the conclusion, when confronted with one of the comparisons and his group response, he stated: *There were so many against me that I thought I must be wrong. Toward the end I got mixed up. I thought I heard the directions wrong. I was beginning to become confused and was more prone to their influence.*

This subject, age 20, yielded eight times. He was active and talked throughout the interview without limit, almost obviating the need for the formal questions. He had yielded consciously on the basis of the judgment that the group was accurate and because of fear of opposing them. He justified his yielding, although he knew better, in terms of unwillingness to spoil the experiment, opening the interview with "Did I foul up your works?" (this referring to his independent judgments!). During the experiment he was preoccupied with "figuring out what to do, what was wrong"; the sense of disappointment with himself came only later. He discovered the purpose of the experiment as soon as the interview opened. Suspicion had flashed through his mind during the session but was not acted upon. His occasional disagreements with the majority seemed to function (*a*) as protest or self-assertion which he did not have the strength to sustain, and (*b*) as attempts to get others to break away with him (on the hypothesis that they were following the leader), which he abandoned when he received no support.

Would you please describe in your own words your experiences during this experiment, that is, what happened, and what thoughts you had? *To be frank, I disagreed with them always after the first three or four trials. I agreed but didn't really agree. I was afraid I fouled up your statistics. Sure I was right, but seven of them had to be right so I gave their answer except that I put in my own opinion every once in a while. Were they always right? I tried looking away, concentrating less or more on drawing imaginary lines—always disagreed with them. I was frustrated because I couldn't agree. I now suspect it was fixed; I came last, you told them the answers. First, the thought flashed through my mind during the experiment, but then I was mostly busy figuring why we didn't agree. I realized you'd know I'd been fibbing so thought I had better tell you and not spoil the statistics. Also, I was thinking the others were following the first fellow; I gave my own answers a few times in hopes they'd join me, was determined to ask you to let me see the cards later, swore I was right. "Fixed" idea didn't really strike me till afterwards. During the* *experiment I assumed they were right and I wrong, and tried to figure out why. Later it occurred to me maybe it was "fixed."*

Did the effect of the others change in any way in the course of the experiment? *No, the level of shock and surprise at disagreement remained constant throughout. Am wondering now (in case it was fixed) what they think of me for following. Wish I'd had the guts. But I went with them, not only because I was sure I was wrong and didn't want to be the only one disagreeing, but because I was sure I was wrong and didn't want to foul up your statistics. I realize that's unscientific, but that is the way it seemed. Then sometimes I thought it worse for the statistics* [to go with the majority]; *so I gave my own answer. Then it would switch. For the last eight or ten* [trials], *I resigned myself to give their answers and see you afterwards to find out what was wrong.*

Did your doubt change as the experiment progressed? *Yes, at first I thought I was right, then I became convinced the other seven couldn't be wrong. I stayed steady then. Toward the end I lost hope they'd come back to me and it was too much for coincidence. Became sure I was wrong.*

Did you have a great deal of doubt at that time? *Visually sure I was right, judged that I certainly was wrong. There were seven others—knew they were all normal.*

Would you say you were disturbed about the disagreements? *Didn't have time to be. It bothered me. I would have been more worried if I wasn't sure I could find out at end of experiment what was wrong. I was too busy wondering to be too disturbed.*

Yielding at the level of action. Included in this group were individuals who yielded although they saw the relations clearly and did not allow themselves to become confused about them. Their compliance was not based on the assumption that the group judgment was superior. Indeed, they skirted the question of truth because another side of the problem became paramount. They were dominated by their exclusion from the group which they took to be a reflection on themselves. Essentially they were unable to face a conflict which threatened, in some undefined way, to expose a deficiency in themselves. They were consequently trying desperately to merge in the group in order not to appear peculiar.

Those who responded in this manner were exposing themselves to the most painful possibilities of the situation. Their compliance had for them a quality of deliberateness, since they knew that they were suppressing their judgments and were unable to mitigate the conflict by becoming unclear. Although they had "decided" to ignore questions of fact, the knowledge that they were acting in this way weighed upon them and made it peculiarly difficult to acknowledge it in a detached way either to themselves or subsequently to the experimenter.

In this group one observes a strange conjunction of open-eyedness and blindness. The subjects embark on the path of compliance in an effort to allay the immediate pain but without a realization of the consequences soon to follow. When these do become apparent they are no longer free to change direction. Although yielding almost without exception, these subjects will also state that "after I became a conformist I was mad at myself," that they "felt better" on the few occasions of independence, and that "it required more of an effort to answer as the group." One obtains the impression of a violently unreasoning persistence, of impulse moving ahead without a cognitive path prepared for it.

This subject is a classic example of the reaction we are trying to describe. He showed pronounced conflict, feeling that his judgments were right but unable to expose himself to the group. He erred eight times.

How many times did you disagree with the group? *Three or four.*

When you gave an estimate that disagreed with the others, did you feel that your judgment was objectively correct or incorrect? *Correct.*

Did you feel that the judgments of the others were objectively correct? *No.*

What did you think when you first heard the others give an answer different from the one you were going to give? *It could be me.*

Did you examine the lines more closely? *Yes.*

Did the others arouse doubt concerning the correctness of your judgments at the time you were giving your estimates? *Skepticism more than doubt. Because there were ten other people disagreeing.* [He describes his doubt as moderate.]

Did you have doubts that your eyes might be deceiving you? *Yes.*

Did this doubt change as the experiment progressed? *Increased, because those looked awful right to me.*

Were you concerned about the disagreements? *Not particularly.*

Try to describe exactly why you were concerned. *That I might be alienating a few people. Here was a group; they had a definite idea; my idea disagreed; this might arouse anger.*

How did you feel when the others and you agreed—when they gave the same answer that you had in mind? *Fine.*

Did you feel that the answers of the others were arousing a tendency in you to agree with them? *Yes.*

Would you say that the pressure was very strong, considerable, moderate or slight? *Very strong. I had a feeling that I was going against their wishes.*

Were you tempted at times to answer as the others did? *Yes . . . we all want to be with the bandwagon.*

As you continued to look at a pair of cards, did the answers of the others continue to appear wrong, or did they seem to become more plausible? *Continued to appear wrong.*

Did you ever give an answer which you knew to be wrong? *Yes.*

How often? *About three or four times.*

Why did you follow the group in these cases? *First, because I figured I probably was wrong, and second, I didn't want particularly to make a fool of myself.*

Did you not feel that you should answer correctly? *When you're in a crowd, who's supposed to know?*

How did you feel when you saw the lines one way and answered in another way? *I felt I was definitely right, even more so when I gave a different answer.*

Did you think the group would disapprove of you, or think you were peculiar if you gave a different answer? *They might think I was peculiar.*

When you gave such an answer, did you at times forget about the question of accuracy, and consider rather that it might make you appear different from the others? *Yes.*

What did you think our purpose was in this experiment? *An experiment.*

As the experiment progressed, did its meaning change? *I began to feel I was being pushed to an answer I didn't want to give.*

At the beginning, after the first three or four judgments, were you aware that you were in disagreement with all the students? *Yes.*

Were you aware that all the students were in agreement with each other? *Yes.*

Did the fact that the others agreed with each other make them appear as a group? *Certainly did —they were trying to put their ideas on me.*

Did they as a result appear closer to each other than to you? *Yes.*

Did they seem to become more united as the experiment progressed? *They seemed united all the way through—only it was more noticeable.*

Did the disagreements produce in you the feeling that you were separated from the group? *Yes and no.*

In opposition to the group? *They were in opposition to me.*

Did it require an effort to answer differently from the others? *No, it required more of an effort to answer the same.*

Suppose I told you that your judgments were always right when you disagreed with the group, what would you think? *A combination of surprise mixed with something of I-told-you-so.* After the full disclosure, he appeared serious, simply commenting: *It's pretty effective.* Invited to explain his reactions, he continued: *I was torn between them and me. At first I was more inclined to disagree with them. As it became more obvious that the disagreements would continue, I changed over.*

We then proceeded to put the following questions:

Were you self-conscious at the beginning of the experiment? *Yes.*

Did you become more self-conscious as the experiment progressed? *Yes, I was standing out as a sore thumb.*

Did you doubt the accuracy of your perception? *At times, yes, not at other times. Sometimes I had a violent reaction—that they were wrong.*

What was your major concern at this time? *To merge into the group.*

Did you suspect that the group intended to affect your judgments? *Yes.*

When did you first begin to feel suspicion? *After five or six* [trials].

Did you retain this suspicion? *Yes. I was very suspicious.*[23] *They were strangers. It was like being in a strange country. "When in Rome you do as the Romans do."*

What in your character and experience would you say was responsible for the way you acted in this experiment? *They are connected—yes. I tend to do things sometimes a little differently from others—jumping at what I can get.* [Here the subject assumes that he was largely independent.]

At the conclusion, in the course of a general conversation, he expressed the view that *the duty of a government is to do the will of the majority, even if you are convinced they are wrong.* Suppose it concerned a lynching? *I wouldn't want to stand in the way, I've seen one. It's like a tide—they'd trample me over—I was run over.*

---

[23] This was an instance of hypothetical suspicion. The subject continued to yield to the end.

The following subject, with ten errors, resembles the one preceding, except for possessing a serious conviction that his action was wrong and called for improvement.

Did the fact that other people were present and attending to the same task affect you in any way? *If people had been alone, there wouldn't be so much agreement.*

How would you describe the effect? *Some people followed. I think I followed a couple of times. I think it was a great deal a group response. If I wasn't absolutely positive, I changed, thinking maybe I'm wrong.*

When you gave an estimate that disagreed with the others, did you feel that your perception was objectively correct or incorrect? *Just the way I saw it. People perceive things differently. Numbers don't make it right.*

Did you feel that the perceptions of the others were objectively correct? *Yes.*

What did you think when you first heard the others given an answer different from the one you were going to give? *I just thought, "to hell with it, I'll give what I think."*

Did you examine the lines more closely? *Yes, I checked it and still thought the same thing.*

Did the others arouse doubt concerning the correctness of your judgments at the time you were giving your estimates? *Yes* (moderate doubt). *I thought that maybe because I wore glasses there was some defect.*

Were you concerned about the disagreements? *No.*

Did you feel that the answers of the others were arousing a tendency in you to agree with them? *Yes—mob psychology.*

Were you tempted at times to answer as the others did? *Yes. The more uniformity among the group, the more it made me conform.*

Were there times when you saw the lines one way and you answered in another way? *Yeah, when I decided to be a conformist on a couple of occasions.*

How often? *Three to four times.*

Why did you follow the group in these cases? Try to describe exactly. *I felt the need to conform. Force of habit—mob psychology builds up on you.*

How did you feel when you saw the lines one way and answered in another way? *Oh—hell,* (smiles embarrassedly)—*I'd see one, then they'd answer, and I'd figure it wasn't important enough to stand out and resist. Do you always get such uniformity?*

Did you think the group would disapprove of you, or think you were peculiar if you gave a different answer? *I didn't at first, but maybe near the end this did operate.*

When you gave an answer that did not agree with your own perception, did you feel that it was objectively correct? *No.*

When you gave such an answer, did you at times forget about the question of accuracy, and consider rather than it might make you appear different from the others? *Yes.*

When in the experiment did this happen? *Anywhere after the fifth* [trial].

What did you think our purpose was in this experiment? *Psychological response. I thought there was something else than eyes.*

As the experiment progressed, did its meaning change? *Toward the middle, I thought maybe it had something to do with pressure.*

Did this situation acquire any personal importance for you? *No.*

This subject's reaction to the disclosure appears notable for its lack of perceptiveness. To the question, "Suppose I told you that your judgments were always right when you disagreed with the group, what would you think?" he replied: "It would inflate my ego." It is understandable that this information, validating his judgment, would produce relief. That he appeared to lose sight of his yielding, responding mildly to the full disclosure, treating the information as a "big joke," and showing predominantly relief and happiness, takes on, in the light of his later statements, a defensive character. Actually this subject was able to extract something positive from his experience. When, toward the conclusion, he spoke more freely, he mentioned that he had always suffered from a strong sense of inferiority, and that he would have complied even more when younger. As it is, he felt that he had overcome this difficulty to some extent and hoped to improve in this respect, predicting that in the future he would be much more independent. Thus he openly recognized his conformity as a weakness which he was working to eliminate.

The following record illustrates the inner turmoil, the fluctuating efforts at rationalization, the contradictions and disturbances that were the consequences of yielding. This subject erred ten times.

Early in the interview he stated: *I must have a difficult time judging lines—I kept getting more and more unsure and gave their answer. Most of their answers* *were faster than mine, much to my chagrin. Why disagree? I couldn't see it too well—a few I just didn't bother with. If seven out of eight are in perfect agreement—nothing to worry about.* He continued: *I was more or less in doubt—wasn't too sure. I can't think they were all so sure. They aren't that much more keen. They were all reacting pretty much—just going along. Why don't you give it to them again—you'd probably get different answers.*

Despite his uncertainties he felt, when he responded independently, that he was right and the group wrong. Also significant was the statement that he "felt better" when opposing the majority.

He further explained the pressure as follows: *From the start there was a tendency to agree with them. I thought they would see them the way I did—they all got them right away—I wasn't too sure of myself then. I was annoyed that I was seeing them differently. I didn't get a chance to make up my mind. Half the time I thought I was wrong and might as well go along. . . . A number of times I thought they were wrong, but I didn't get a chance to think it out myself. I squirmed through the whole thing, thinking I'll just go along with them. . . . It was half and half—I wasn't so sure—felt their answers must be right, and I would be peculiar . . . I agreed because I wanted to agree with them. Takes a lot of nerve to go in opposition to them.* At times he mentioned that "I didn't even bother to look." He also upbraided himself for not being "more assertive."

Toward the conclusion he volunteered a further example of his group consciousness. Around election time he marked Dewey on a student ballot, although he preferred Truman, because he thought Dewey would win and that he was preferred by most. He was making the point that he went along with the majority even when he could have acted on his own belief.[24]

The following protocols are of interest mainly for the light they throw on the elaborations and justifications for compliance when it becomes a settled mode of the individual's conduct, which he is unable to accept outright or to discard.

This was a sophisticated young man of eighteen with a mock-heroic swagger, who responded with defensively pretentious generalizations from his social science courses. He erred 10 times. He had the air of inviting one to join him in a subtle joke which of course he alone could understand. He was also nervously ready to burlesque himself if his swagger failed to impress as a "straight" performance. Before admitting to yielding, of which he was aware, he fenced as long as possible. As far as we could see, he stopped trying to decide which answer was right, taking refuge in a feeling of amusement while following the group.

His feeling of amusement functioned in a way that reveals the core of one kind of cynicism. He considered that the others were following the leader, and so he went along with a contemptuous, amused feeling that he was nobody's fool and that nothing was being put over on him. When asked why under these conditions he joined the farce instead of exposing or resisting it, he readily explained (and this he saw as a still deeper joke in the situation) that he was not certain it was a farce and therefore he protected himself from exposure to (remotely possible) wrongness. His attitude protected him from the humiliation of yielding, but was at the same time necessitated by lack of courage.

Would you please describe in your own words your experiences during this experiment? *Frankly I thought the mob were following the first man. Of course, it could be my eyes since I was the only one who disagreed.*

Try to think back to the time that you were giving the judgments. Did the presence of the others affect you in any way? *Yes, people tend to follow the leader. When I disagreed I felt outside the group.*

Still thinking back to the time that you were giving your judgments, would you say that the others made you doubtful about your accuracy? *Yes. Being in a minority might have a tendency—it might have had a tendency to make me change my mind, but it didn't.*

How did you feel when you continued to give answers different from the others? *There's naturally a tendency to wonder what the others are thinking. Really couldn't say what I thought.*

---

[24] These comments raise a further problem concerning the operation of group pressure that we have not studied but which deserves investigation. It appears that compliance with a group trend can be strong although one suspects that the group itself is misguided and conformist.

Did you ever answer as the others did, against your own first choice? [Gay and subtle] *Yes. You don't make your own choice, you let others choose for you. Instead of proving you're wrong, you let yourself be right.*

On these occasions, as you continued to look at a pair of cards, did the answers of the others continue to look wrong, or did they begin to look more right? *Wrong. Just doubting your own abilities. There's the old tendency there not to mind being wrong in company. Henry Adams wrote about that.*

What did you think about the other people in the group when they all gave an answer that looked wrong to you? *Either they all saw the same thing and I was wrong, or they were following the first guy. Thought the experiment would be just as good if only the first guy were there.*

Did you wonder what everybody might be thinking of you when you disagreed? *Yes, I had a tendency to wonder what others are thinking about when you get out of the groove.*

Suppose I told you that your judgments were always right when you disagreed with the group, what would you think? *Wouldn't break any buttons over it.*

Following the full disclosure he stated: *I even had a hunch of that, a stray hunch, never serious.* He then continued: *If it had been on some other issue, I would have stood my ground. But on a question of faculties, like this, you get unsure. When you've tried once or twice and you're still wrong, there is a tendency to crawl back into the fold. I also felt it was a game of follow the leader. Felt amused.*

How did you feel when you saw the lines one way and answered in another way? *Frankly, I was thinking over in my mind how people will follow the rule of the majority.*

The following subject, age 20, reveals a reaction that cannot be readily included among the preceding categories. He yielded eight times. Although the situation was quite salient for him, and his degree of disturbance considerable, he was unaware of the impact of his yielding and rationalized it in "philosophical" terms.

Would you please describe in your own words your experiences during this experiment, that is, what happened, and what thoughts you had? *There was general agreement at first. Around the middle, I started doubting which was which. I said what I saw, but guess I was wrong when I disagreed. Majority is usually right.*

Did the presence of the others affect you in any way? *Yes. Toward the end, I felt slightly disturbed, not confident.*

Did the effect of the others change in any way in the course of the experiment? *Usually I agreed, sometimes felt they were dead wrong, but I guess they weren't. If I had any doubts whatsoever, I'd perfectly naturally go with the majority. I'm a political science major; I carry those principles over into psychology.*

Did you feel that the comparisons during the later part of the experiment were easier or more difficult than in the early part? *Obviously more difficult, no doubt. They seemed that way, anyway.*

Still thinking back to the time that you were giving your judgments, would you say that the others made you doubtful about your accuracy? *Yes. Who am I to disagree with everybody? Sometimes I did, of course. I'm perfectly willing to stick my neck out. My thought was: "I can't subvert my principles that far." I won't agree just for the sake of agreeing, but will go along when in doubt.*

Did you ever think that your eyes might be deceiving you? *Yes. Suppose my eyes got tired. I was quite sure in my own mind I was definitely wrong.*

How did you feel when you continued to give answers different from the others? *Doesn't bother me. Would just as soon disagree, usually do.*

Would you say you were disturbed about the disagreements? *Not a bit. Took it rather easily.*

Would you say that you were tempted at times to answer as the others did? *Yes. If there's an equal balance between two alternatives, I think it natural and right to permit the majority to influence your answer.*

Did you decide at any point to agree with the group, or to be independent of them? *No. I let every case be separate. Not a fair test if I didn't.*

What did you think when the others gave the same answer you had in mind? *They're right, and I'm right, too.*

How do you explain it? *Sometimes I felt they all were going along with the first guy who answered. Not much strength of mind.*

What did you think about the other people in the group when they all gave an answer that looked wrong to you? *Most people are sheep anyhow. Sometimes I thought they seemed to be wrong, but probably I am.*

While you were comparing the lines, what did you think our purpose was in this experiment? *If there was disagreement, why? And whether the group behavior influenced results. Am sure if no one were there I would have been equally doubtful.*

What is your opinion of psychology experiments generally? *They have contributed great deal to our knowledge. Wouldn't go into it myself, I'm not that interested in people. I'm interested in government, how to make people work together. If we must have government, I'd like to be in it.*

Suppose I told you that your judgments were always right when you disagreed with the group, what would you think? *Wouldn't believe you. I think a majority is usually right.*

After the full disclosure he laughed hard: *A dirty trick! Apparently toward the end I was really influenced. Said what I thought right most of the time.*

Did you suspect that the group intended to affect your judgments? *No.*

This experiment lasted a fairly long time. Could you tell us how your feelings and attitudes changed as the experiment continued? *When you begin to disagree there is some doubt. When doubt continues against insuperable odds, there is a strong tendency to say "what the hell" and go along.*

Did the situation in general become easier or more trying? *Never trying. After disagreeing for so long, you figure you might as well agree, especially since it isn't easy to judge and there's some doubt.*

Some of the preceding protocols contain a sense of resentment at the majority. It is not surprising to find this reaction to compliance that violates conviction. It also suggests, what our observations otherwise confirm, that yielding was not a mark of respect for the majority or a sign of closer feeling with it.

D. *The Question of Right and Wrong*

Among the questions facing the subject, one of the most crucial was whether he or the majority was judging veridically. Let us attempt to see whether there was an orderly relation between performance and the answers this question received. Among the possible alternatives the following were the most usual: (*a*) The majority is wrong, I am right; (*b*) I am right; (*c*) I am wrong, the majority is right.

We will consider first the independent subjects. (*a*) Included among the independents are those who arrived at the first conclusion. As we have seen, the emotional reactions of this subgroup were not homogeneous. (*b*) Also included among the independents were those who adopted the second conclusion, namely, that they were right, while refusing or reluctant to commit themselves about the majority. (*c*) The third conclusion was also compatible with independent performance (see pp. 40–41). It is clear that the answer to the question "who is right, and who is wrong?" did not automatically decide whether one was going to be independent. The first two alternatives pointed toward independence, and the last toward yielding, but independence was compatible with the entire range of alternatives. Supervening upon these alternatives were subsequent emotional reactions which could take different directions and produce divergent decisions.

Turning to those who erred it might appear that there were none who adopted the first or second conclusion, and that yielding was compatible only with the third alternative. But there are grounds for holding that all the alternatives were also represented among those who erred.

The nub of the difficulty concerns the meanings of the terms right and wrong in the present context. "The majority is right" can mean that they are judging the given data correctly; it is this comparatively specific sense that figured in the thinking of independent subjects. But yielding subjects extended the assertion to mean that the majority was generally right. They endowed the majority with a quality of rightness or authority that went beyond the immediate question at issue, granting it a general and undefined superiority that was capable of overriding their convictions about their own rightness in the specific case. This assumption of group superiority expressed itself, however, in diverse ways. 1. Some yielding subjects rested their case ostensibly on a simple probabilistic inference. But when they did so they were also denying to their judgments the status of evidence, trying to see the relations as the majority apparently did, and at times allowing themselves to become confused. 2. A deeper effect is observed in those who followed the majority because they found open opposition to it unendurable. These subjects continued to see the relations clearly and even appeared no longer concerned with questions of accuracy. They granted the majority a pervasive and more absolute kind of legitimacy; the majority wielded the sinister power of rightness although its actions were distorted—because it was the majority. 3. We can say least about the few who complied without awareness (see pp. 42–43). Possibly these individuals acted on the assumptions of the preceding groups. It seems justified to conclude that yielding subjects granted superior credence and power to the group, but that the pressure took effect in different ways.

From what has been said it follows that yielding was incompatible with the first alternative when taken in its specific meaning. To state the point differently, yielding could not have the standing of a "Galilean" act. The absence of coercion excluded the defiant suppression of one's experience in order to escape ignorant or cruel group sanctions.

E. The Problem of Individual Differences

It will be evident that we have described the features, rather than explained the grounds, of independence and yielding. Our procedure was not designed to throw light on this question. The interview was concerned mainly with the overt content, and the subjects could not be expected to inform us of the functional determinants of independence and its failures. The striking individual differences leave us therefore with an unsolved problem that requires investigation. At this point we will make only some first comments.

There are two directions open for an interpretation of the observed individual differences. They may be the product of momentary, episodic circumstances. Or they may be a function of consistent personal qualities which individuals bring into the experimental situation and which have as decisive a bearing on their actions as any of the other conditions we have studied.

The contingency thesis might be formulated as follows. Although the experimental setting was in its general contours similar for all subjects there is no doubt that it was not completely controlled. The composition of the majorities varied from time to time. Even when they remained constant a majority could hardly duplicate its actions exactly. Doubtless the majority also responded in subtly different ways to each subject. Nor did we have control over the subject's immediate reaction to the group or over his definition of the situation which may have depended on his momentary condition of gaiety or preoccupation. These and numerous circumstances of this kind might be responsible for the observed variability. According to this view, the same individual who went consistently with the majority might with equal likelihood have been independent if the circumstances described had varied in slight ways.

Although such contingent conditions probably played a part, there are strong grounds for holding that they accounted for the results only partially. The evidence cited in this chapter, particularly that pertaining to the different modes of reaction, produces the insistent conviction that the responses to the experimental conditions were functionally connected with crucially important character-qualities. One could not observe the subjects and escape the impression of a coherence between their actions and the kind of persons they appeared to be. While these observations and impressions cannot take the place of evidence they may be the starting point for thinking.

To proceed with this problem it would be most desirable to establish the degree of individual consistency of independence. The most convincing way would be to observe the same person in different situations each of which poses an alternative between independence and yielding (while excluding the subject's conscious comparison between them).[25] Or one might formulate relevant hypotheses about the relation between the experimental performance and certain character qualities, the presence of which would be established independently. The combination of both procedures would be most pertinent.

Concerning the probable grounds of independence and compliance as characteristic modes of coping with social opposition we can here only say some preliminary words. In the first place the problem needs to be specified. The issue concerns the probable reasons for differences in tolerance of the denial of one's private experience by others. To face such denial is for most who have had a social past an ordeal. Under this stress some adhere to their direct experience despite doubts and pains, while others abandon its claims. We are concerned with the reasons for differences in endurance of this difficulty.

Social opposition of the kind here referred to is an ordeal because it exposes the individual to misunderstanding, calumny, or ostracism. Differences in endurance are therefore differences in capacity to withstand social misunderstanding or suspicion.

Our observations suggest that independence requires the capacity to accept the fact of opposition without a lowered sense of personal worth. The independent person has to organize his overt actions on the basis of experience for which he finds no support; this he can do only if he respects his experiences and is capable of claiming respect for them. The compliant person cannot face this ordeal because he translates social opposition into a reflection of his personal worth. Because he does so the social conflict plunges him into pervasive and incapacitating doubt.

---

[25] It would be enlightening to establish whether there is a relation between performance under the present conditions and those resembling the classical suggestion experiments in which the source of influence is largely or entirely unknown to the subject. Within the range of the present procedure it should be possible to study the effect of a majority when its deviations from accuracy are sufficiently small to prevent the experience of outright conflict and to compare the performances of individuals with those obtained under the present conditions.

A guiding assumption concerning the sources of these differences is that the experimental dilemma communicates in an important way with central regions of the individual's forgotten past. It communicates with deep-seated feelings of inadequacy, isolation, of real or imagined deficiency, as well as with capacities for mobilizing resistance to these. A further and more specific assumption is that psychic inadequacies of this order are in part a consequence of the denial by others of the worth of one's experiences, and that one of the consequences of such denial is self-depreciation and estrangement from one's own experiences. Few grow up without some denial of their feelings and insights by others. The child who is assured that he is attached to his infant brother, when other emotions toward his sibling predominate, or that his parents are devoted when sees them bursting with anger, is faced with feelings that have no validation in the world of others. If feelings of inadequacy are present they are likely to blunt reliance on one's own observations and breed confusion. Aside from such experiences, which resemble in a measure the structure of the present experimental conditions, there are the daily, unremitting pressures to conformity which shape not only the practices of persons but also their most intimate conceptions of themselves. The pursuit of this problem therefore brings us to the uncharted area of the relation of the person to his own experiences and convictions, and of the relation between these and interpersonal ties. It is in the clarification of these personal-emotional processes that the solution of the present problem lies.

A point that may complicate the clarification of the problem concerns the observed differences in responsiveness to the majority. If we could assume that all subjects faced the same difficulty it would follow that performance was a function of ability to cope with it. This was not the case; some experienced a more severe conflict than others to the opposition of the majority (see pp. 35–36). The resulting performance was therefore a function both of the cogency of this kind of conflict for the individual and of his ability to overcome it. Some independent subjects were less sensitive than others to the import of the disagreement, without being necessarily more confident than others in their own judgment. Some yielding subjects experienced a severe conflict, while others did not appear much shaken, did not feel deeply about yielding, and the requirement to report truthfully was apparently not a very strong motive with them. The latter may have yielded for the reasons mentioned rather than for lack of confidence. These considerations leave open the possibility that the differences between independent and yielding performances were in part due to the severity of the conflict rather than to differences in capacity to withstand the same degree of conflict.

We must leave open the question whether and how constitutional differences contribute to the observed individual differences. It is conceivable that there are systematic differences between persons in capacity to endure tension or in ability to cope with contradictory forces, to weight their respective merits without becoming overwhelmed.

The reader who has followed the account to this point may have wondered whether this mode of experimentation, which places an individual in a situation that can cause pain and embarrassment, is justified, and whether its effects are desirable. We will attempt to state briefly our thinking concerning this question as it developed in the course of these experiments.

On the basis of our observations and of the comments they made, we are of the view that the greater number of subjects felt that they had taken part in an experiment with a bearing on serious human questions, and one that justified the brief conflict to which they were exposed. They found it illuminating to consider their earlier doubts, temptations, and surrenders in the light of the subsequent disclosure of the actual state of affairs. The experiment had provided a vivid illustration of the importance of individual independence and of the dangers of certain forms of group pressure; in many cases it aroused wonder and reflection about the role of these forces in social and personal life.

Their conduct also confronted the subjects in the end with certain evidence about themselves that they used variously. If they were open to self-knowledge the disclosure furnished them an opportunity to evaluate their conduct and to appreciate more clearly certain trends in their character. When they had been independent they found encouragement in the capacity they showed to meet the challenge of the experimental episode and were somewhat wiser about the weaknesses they had to battle. Those who had yielded to the majority were often able to show more respect for the feelings of resistance and the struggle for independence that all in some measure displayed. (This may be one explanation for the focusing of many yielding subjects on their occasional independence.) Others could not use the knowledge of their shortcomings constructively; these slurred over the significance of the situation and refused to think about it. It would be too much to expect that a brief episode in a permissive setting which touches off conduct that has a long history, would produce lasting changes for good or ill. We are inclined to hold that where further changes did occur they were in a favorable direction and that those who found it difficult to defend their conduct acted so as to obscure the facts and the problems they raised. As far as we were able to observe there were no harmful effects.

Assuming the correctness of the preceding observations the issue is by no means completely resolved. Unfortunately the problem we studied excluded the possibility of obtaining the subject's permission in advance. To be sure, the subject was not placed in a situation in which he was forced to suffer indignity; he could not fail to see that his conduct was of his own making. The present procedure thus had nothing in common with those which, for some experimental purpose, insult or embarrass the subject. Nevertheless the involuntary participation, the fact that the subject was, without his knowledge or consent, placed in a situation in which he revealed himself (without the other participants revealing themselves in turn) remains a problem.

To justify this mode of experimentation it is necessary to take the position that persons engaged in the normal occupations of life and shouldering its usual responsibilities should be considered capable of handling truth emotionally, that we have no moral obligation to shield them from everyday facts about themselves. To take part in social life implies the ability to meet new experiences: either by profiting from them or by shutting them out if they appear too threatening. (We do not here consider whether it may not be our obligation to challenge each other to bear more truth.)

Nevertheless the circumstances place a special responsibility on the experimenter and obligate him to surround the procedure with proper safeguards. It has been the writer's experience that far more important than the momentary pain or discomfort of the procedure is the way in which the experimenter deals with the subject. When subjects grasp the aim, when they see that the experimenter respects their feelings and that he depends upon them for help in clarifying what has happened, they become for the time being collaborators and cease to regard the problem from an entirely personal standpoint.

V. Variations of the Experimental Conditions

The present chapter describes a number of variations the aim of which was to clear up certain questions of interpretation arising out of the preceding experiment (Experiment 1). The main feature of the earlier procedure was retained: there was always a unanimous majority of 7 to 9 persons and a naive minority of one. As in Experiment 1, the subjects (and the members of the majority) were male college students.

Experiment 2: Varying the Modality of Judgment

Nothing in the observations of Experiment 1 suggests that the majority effect was peculiarly a function of the dimension of length which we employed because of its convenience. It seemed nevertheless advisable, in view of our limited knowledge, to observe whether the effect could be obtained with an entirely different continuum. Accordingly the present variation was performed, the task being to compare the brightness of two discs.

The subjects viewed two color wheels, of which the one at the left was the standard. They were instructed to judge whether the (comparison) wheel at the right was lighter, darker, or of the same brightness as the standard. Between trials a screen was placed in front of the wheels while the experimenter adjusted the one at the right and checked to see that there was no flicker.

The standard color wheel was throughout set at 90° W+270° Bl. In the course of the trials the comparison wheel was set to the following 6 degrees of white: 10°, 30°, 60°, 130°, 180°, and 260°. In short, there were 3 steps above and 3 steps below the 90 ° W of the standard. According to the observation of the experimenters the steps were easily distinguishable from each other and from the standard. The points were chosen so that they would be at approximately equal intervals. The comparison disc was never set equal to the standard.

The color wheels were 7 1/4 inches in diameter and the distance between their centers was 47 inches. The subject sat directly in front of the wheels and at approximately the same distance from them as those in Experiment 1 were from the lines. He did not know that only one wheel was adjusted. The speed of rotation was the lowest at which no flicker was visible.

On critical trials the majority judged wrongly, calling the darker comparison disc brighter, and conversely. Announcements of the majority were always in terms of "darker" and "lighter." Consequently judgments of "equal" by minority subjects are to be considered as compromise reactions.

There was a total of 23 trials of which the first 5 were for the purpose of accustoming the subjects to the procedure. Trials 1 and 2 were neutrals; the settings of the comparison wheel were extreme so that the judgments were very obvious. Trials 3 and 4 were critical "buffer trials," the discrepancies being small, 10° in each case; Trial 5 was another neutral trial. There followed 18 trials (6 to 23) with which we are here concerned and which are summarized in Table 14. They, consisted of 12 critical and 6 neutral trials, distributed in a random order. Each critical setting was presented twice.

TABLE 14

BRIGHTNESS COMPARISONS AND MAJORITY RESPONSES: EXPERIMENT 2

Trial	Degrees of white of comparison color wheel	Correct response	Majority response
6	60	D	L
7	180	L	D
8	130	L	D
9	10	D	D
10	30	D	L
11	260	L	L
12	130	L	D
13	30	D	D
14	10	D	L
15	60	D	L
16	180	L	L
17	260	L	D
18	30	D	L
19	130	L	L
20	10	D	L
21	60	D	D
22	260	L	D
23	180	L	D

Note.—The preliminary five trials, which were introduced for purposes of explanation, are not included. L = Lighter; D = Darker.

There were 11 subjects in this variation. The errors of the subjects were, in order of increasing frequency, as follows: 0, 0, 0, 1, 2, 3, 3, 8, 11, 11, and 12. Eight of the 11 subjects showed one or more errors. The range of individual differences was great, from complete independence to complete yielding. The mean number of errors was 4.6 or 39 per cent of all critical estimates. Five out of a total of 51 errors were compromises.

In Table 15 the frequency of errors for each of the steps is tabulated. Because of the small number of intermediate subjects the relation of errors to the magnitude of discrepancy remains undecided.

The main conclusion we draw is that

TABLE 15

FREQUENCY OF ERRORS ON BRIGHTNESS COMPARISONS: EXPERIMENT 2

Setting of comparison wheel (in degrees of white)	10°	30°	60°	130°	180°	260°
Number of errors	8	8	7	13	9	6

the majority effect was present under these conditions. Further, our observations of the subjects and of their reactions during the postexperimental interview leave no doubt that they were independent or acquiescent for essentially the reasons that obtained in Experiment 1.

The level of tension appeared somewhat lower than in the main experiment, but the observations were not continued far enough to clarify the reasons.

We conclude that in its main aspects the present situation was psychologically similar to that of Experiment 1, and that the majority effect there observed is not restricted to a particular sensory dimension.[26]

Experiment 3: A Further Variation of the Dimension of Judgment

The stimuli of Experiment 1 were lines 3/8 of an inch wide. It might be contended that since the lengths were actually rectangles, the judgments involved the dimension of form and not of length alone. Although our observations did not support this interpretation, it seemed desirable to limit the procedure more strictly to the comparison of lengths.

Accordingly the experiment was repeated with lines 1/32 of an inch wide. The trials and other details of procedure were those of Experiment 1. Following the eighteenth trial with the thin lines the subjects were given the first 9 trials of Table 1, this time in the original width of 3/8 of an inch. There were 11 subjects in this variation. The results are summarized in Table 16.

Errors occurred with somewhat lower frequency than in Experiment 1, comprising 24.5 per cent of the critical estimates. Extremely yielding scores were curtailed, but the number of cases is too small to establish whether this difference is significant. The difference between the proportion of errors in this variation and in Experiment 1 is not significant; if we dichotomize the distributions into 0–3 and 4–12 errors, $\chi^2 = .179$.

When the stimuli of Experiment 1 were restored in the final trials the level of errors was much the same as with the thin lines, accounting for 27.4 per cent of the critical estimates. There was also a curtailment of extremely yielding scores. Finally, the individual performances under the two conditions were significantly associated. Comparing the concordance of ranks (see Footnote 9) for the number of errors produced by each subject under

TABLE 16

DISTRIBUTION OF ERRORS: EXPERIMENT 3

Condition	Subjects										
	1	2	3	4	5	6	7	8	9	10	11
Thin lines (12 trials)	0	0	1	1	3	3	3	4	4	6	7
Thick lines (6 trials)	0	0	1	1	0	2	3	1	4	2	4

---

[26] Since we made no effort to equate the brightness differences with the size differences of Experiment 1, we cannot compare the magnitude of the effects in the two cases.

each of the two conditions we obtain an approximation to chi square of 22.7, which with 11 degrees of freedom gives a p between .02 and .01. The rank correlation between errors under the two conditions equals +.75. The subjects showed a definite tendency to maintain to the end the stand they adopted earlier in the experimental session.

We may conclude that the majority effect occurs when the comparisons are strictly restricted to the dimension of length. If we bring these results in relation with those of Experiment 2, we have added basis for the conclusion that the majority effect of the present studies is not limited to a particular sensory dimension.

Experiment 4: The Import of Public Judgments

One distinctive feature of Experiment 1 was the requirement to announce one's estimates in public before a group that was also estimating in public. It stands to reason that this procedure may have contributed to the majority effect. The task of the minority-of-one was not only to arrive at a private decision in the face of opposition; he also had to announce his judgment within the hearing of those who disagreed with him. In the following variation we altered this condition. The majority continued to respond publicly as before, but the critical subject stated his estimates in writing.

The present variation is part of an experiment designed primarily to study a different problem to be reported subsequently in detail; we will therefore state the procedure here only summarily. In brief, we arranged to have the critical subject arrive "late" in the experimental room. The experimenter offered an acceptable reason for not being able to include the subject in the procedure of public announcements, but invited him to take a seat (which happened to be, as in Experiment 1, next to the last) and to put his estimates in writing. In the course of this explanation the experimenter managed to include the full instructions. In addition, the experimenter explained that he was interested in the time relations of the judgments, that he would signal to each member, with the help of a metronome, when to announce his response, and that the critical subject was to wait with writing his estimate until his turn came. This procedure insured that he would respond, as in Experiment 1, only after all but one of the majority. Thereupon the experimenter "continued" with the trials of Experiment 1 (which were numbered, for the sake of the critical subject, from 7 to 24). The majority of course responded publicly. In short, we duplicated the conditions of Experiment 1, except for the mode of response of the critical subject.

There were fourteen subjects in this variation.

The results appear in Table 17. It includes the distribution of errors on critical trials and, for purposes of comparison, the distribution of errors in Experiment 1 and in the control group.

The shift from public to silent judgments markedly lowered the frequency of errors. The mean frequency of errors was 1.50, or 12.5 per cent of all critical estimates, about one-third of the errors found in Experiment 1. The results were intermediate between Experiment 1 and the control condition. In Experiment 1 the errors ranged up to the maximum of 12, in the control group they did not exceed 2, while under the "silent" condition the maximum was 6. Because of the small number of cases in this variation and the great range of errors in Experiment 1, we cannot establish a conclusive difference statistically

TABLE 17

Errors on Critical Trials: Experiment 4

Group	N	Number of Errors												
		0	1	2	3	4	5	6	7	8	9	10	11	12
Experiment 4	14	5	4	1	3			1						
Experiment 1	123	29	8	10	17	6	7	7	4	13	6	6	4	6
Control group	37	35	1	1										

between the two conditions. However, inspection of the data leaves little doubt that the two conditions were producing different effects. The difference with the control group is significant ($t=3.64$, which with 49 degrees of freedom, gives a $p<.01$).

Other aspects of the results confirm the reduced character of the group pressure under the present conditions. We found in Experiment 1 that on extreme trials—trials on which the majority chose the most discrepant comparison line—the critical subjects mostly followed the majority; 81 per cent of the errors were extreme, and 19 per cent were compromise reactions, or errors nearer to the standard. (See Table 5.) With the critical subjects responding silently all errors (of which there were 10) were moderate, no one followed the majority when it became extreme, a result that confirms the diminished power of the present majority.

The difference described holds not only in general but also for each critical trial. Figure 6, which plots the errors on successive critical trials, demonstrates this fact. It also reveals a striking parallelism between the fluctuations of the curves in this experiment and Experiment 1, again confirming the presence of stable differences between the stimulus-comparisons, this time despite a considerable shift in the general level of performance.

Why did the condition of silent judgment fail to abolish completely the occurrence of errors? The statements obtained during the interview provide a strong hint. All but one of the critical subjects assumed that the experimenter would compare their written estimates with the majority's. They added that this presumption had no effect upon their judgments, but this statement has little evidential value. We will have ample occasion to see in later studies that critical subjects grossly underestimate the effects of helpful circumstances upon their independence, that they attribute independence to themselves and not to conditions which we can demonstrate to have altered their actions decisively.

It is clear that the silent judgments were not private in the full sense. This raises a

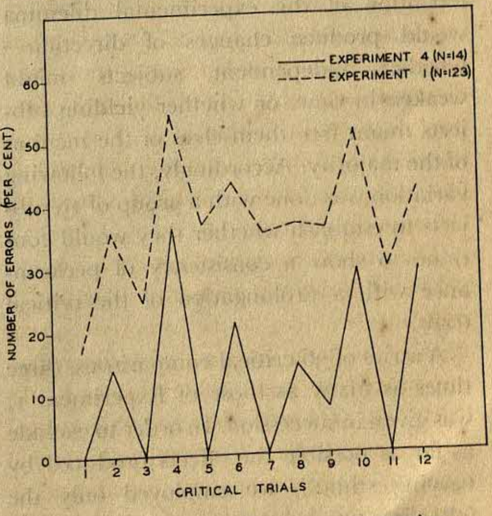


Fig. 6. Frequency of errors on successive critical trials: Experiments 1 and 4.

number of questions that await further investigation. One might attempt, for example, to surround the individual's judgments with more complete privacy in order to establish the sheer effect of a disagreeing majority.

It would also be of interest to exclude the bodily presence of the majority while retaining the knowledge of its disagreement. This could readily be done by merely informing the subject how a group of others judged the materials before he sets down his own judgment. We would then be in a position to compare the effects of a predominantly cognitive awareness of a conflict with those produced by a physically present group.

Experiment 5: Role of the Temporal Factor

We have found that the subjects of Experiment 1 were highly consistent in coping with the majority; early independence was prognostic of later independence, and similarly for yielding. The absence of a cumulative effect of majority pressure is a matter of considerable interest. Since the duration of Experiment 1 was limited, it seemed reasonable to ask whether a prolongation of the experimental dilemma would produce changes of direction—whether independent subjects might weaken in time, or whether yielding subjects might free themselves of the incubus of the majority. Accordingly, the following variation was done with a group of 19 subjects to establish whether they would continue to show a consistency of performance with a prolongation of the critical trials.

A series of 36 critical comparisons, three times as many as those of Experiment 1, was given in succession. In order to exclude as far as possible the effects produced by varying stimuli, we employed only the following two comparisons:

Standard	Comparison lengths			Majority choice
5	5	4	6 1/2	4
6	5	6	7 1/2	7 1/2

Each of these was shown 18 times. On successive trials the comparison lines appeared in different positions. The instructions and other details of procedure were those of Experiment 1. The series opened with 2 neutral trials; there were no neutral trials thereafter in the case of 11 subjects. Eight subjects continued to receive the same pair of neutrals after each set of 6 trials. The results were so very similar that we present the data jointly.

The results closely resemble those of Experiment 1. Errors varied from 0 to 36, and the mean frequency was 37.6 per cent of the total judgments. In a control group of 11 subjects the total number of errors was 4, or 1 per cent of the critical estimates. Of greater interest to us is the consistency of the reactions. For the group as a whole the level of errors is remarkably constant, as is evident in Figure 7, which plots the frequency of errors on successive critical trials. With the exception of the first trial there is no change of direction throughout the series. To establish whether the mass consistency is based on a corresponding consistency of the individual subjects we have obtained their scores in each third of the series (each based on 12 trials). Most subjects were found to be highly consistent throughout. In particular, those who were initially entirely independent or extremely yielding changed least. The only appreciable changes occurred in four initially intermediate subjects, and they changed in both directions. The mean errors on successive thirds for the entire group were 4.05, 4.68, and 4.78, respectively; the differences between them were not significant.[27]

---

[27] Dr. Eugene Galanter has proposed a simple and attractive test of the consistency of individuals

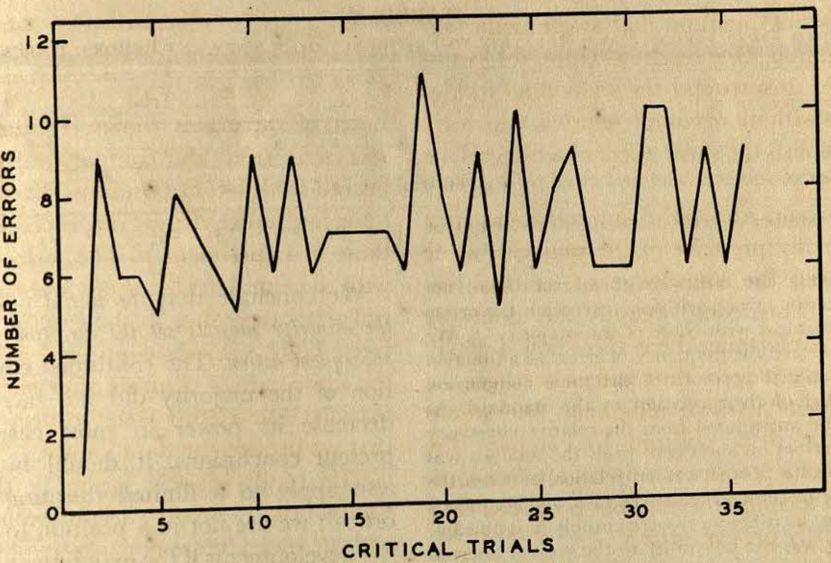


Fig. 7. Frequency of errors on successive critical trials: Experiment 5.

A more detailed analysis of the trials, which differed systematically in certain respects, revealed the following. 1. The frequency of errors was greater on all but the first 8 critical trials with a 6-inch standard than with a 5-inch standard. Of a total of 257 errors, 153 (or 60 per cent) came from the comparisons with a 6-inch standard, this despite the fact that the magnitude of contradiction was larger than with the 5-inch standard. This result is in accordance with the findings of Experiment 1 (see pp. 18–19), and is further substantiated by the results of Experiment 6. 2. With the 6-inch standard the majority judgments were extreme. In accordance with the findings of Ex-

---

which should prove useful when the data are less decisive. Let us assume that on each set of twelve trials the individual has an equal likelihood of making errors between 0 and 12. On the assumption that the errors occur randomly we can compute the probability of differences between the sets as follows. Suppose we have 13 balls in an urn, numbered from 0 to 12. If we make two independent drawings with replacement, and observe the absolute value of the difference, we should find in the long run that most of the differences are small, since there are 13 ways in which a difference of zero can occur, but only 1 way in which a difference of twelve can occur, i.e., if one of the balls is a zero and the other is a twelve. We plot this theoretical probability distribution in Fig. 8. The median difference in this distribution is greater than 3. Therefore we should expect, on the chance assumption, that subjects would show a difference less than 3 about as often as they show a difference greater than 3. The obtained probability distribution with the 57 pairs of observations (19 subjects times three combinations) is also plotted in Fig. 8. We observe that 50 of the 57 observations are less than 3. We would expect an outcome such as this less than once in 1,000 experiments.

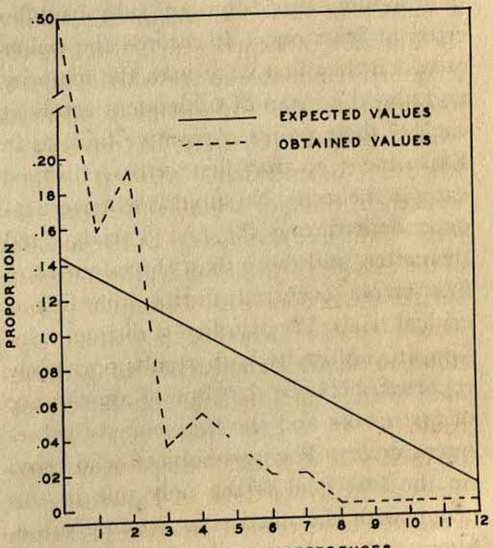


Fig. 8. Expected and obtained distribution of absolute differences: Experiment 5.

TABLE 18

POINT OF FIRST YIELDING AND RELATION TO FREQUENCY OF SUBSEQUENT ERRORS: EXPERIMENT 5

	Trial							
	1	2	3	4	5	6	7	8
Number of subjects who first erred on a given trial	1	8	1	1		1		1
Mean number of errors subsequent to first error	35	20.8	13	25		0		5

periment 1 the errors on these trials were not compromise errors; with one exception the errors were identical with those of the majority. 3. We also analyzed the frequency of errors as a function of the spatial positions of the three comparison lines and of their relation to the standard. As might be anticipated from the relative constancy of the errors on successive trials the analysis was negative. (*a*) There was no relation between the position of the correct alternative and frequency of errors; all differences were entirely insignificant. (*b*) Nor was the proximity to the correct alternative of the standard and lines chosen by the majority of consequence. Again the differences were not significant.

The prolonged trials of this variation permit us to check the previously reported finding concerning the relation of latency of yielding to amount of yielding. (See pp. 21–23.) Table 18 contains the relevant information, based on 13 subjects who erred at least once. It records the points at which they first went with the majority and also the mean of subsequent errors at each of these points. Again we find, as in Experiment 1, that first errors clustered early in the series. No subjects who were independent during the first 8 trials erred thereafter, and more than three-fourths of first errors occurred during the first 2 critical trials. We also find a distinct relation, in conformity with results previously reported, between the time of appearance of first errors and the frequency of subsequent errors. The one subject who erred on the first trial is the only one in this group with the maximum possible subsequent errors. Among the 8 subjects who erred on the second critical trial the minimum number of subsequent errors was 9. The lowest errors are to be found in subjects who started to yield latest.

We conclude that *the initial reactions of the minority subjects set the direction for their subsequent action.* The continued contradiction of the majority did not increase or decrease its power in most cases. The present conclusions, it should be noted, also apply to a limited duration of the ordeal; we are not in a position to foretell the developments if the experimental situation had been prolonged much further.

Experiment 6: The Effects of Size of Standard and of Absolute Discrepancy

In Experiment 1 the incidence of errors was to a remarkable degree a function of the particular stimulus relations. There was evidence of a tendency for errors to increase with size of standard but not with the magnitude of the contradiction. Because many relations varied simultaneously it seemed advisable to study the contribution of these factors under more systematic conditions. This was the purpose of the present variation.

The comparisons were between a standard and a *single* comparison line that was always unequal to the standard. There were 4 standards; their lengths were 3, 5, 7 and 9 inches, respectively. Each standard was paired with comparison lines

TABLE 19

STANDARD AND COMPARISON LINES (IN INCHES): EXPERIMENT 6

Standards	Comparisons					
	−3	−2	−1	1	2	3
A 3″	2.25	2.50	2.75	3.25	3.50	3.75
B 5″	4.25	4.50	4.75	5.25	5.50	5.75
C 7″	6.25	6.50	6.75	7.25	7.50	7.75
D 9″	8.25	8.50	8.75	9.25	9.50	9.75

TABLE 20

ERRORS AS A FUNCTION OF SIZE OF STANDARD AND SIZE OF DISCREPANCY: EXPERIMENT 6
("M" denotes majority errors, "C" denotes compromise errors)

Length of standard	Discrepancy with standard												ΣM	ΣC	Σ(M+C)	Mean	Per Cent
	−.75″		−.50″		−.25″		.25″		.50″		.75″						
	M	C	M	C	M	C	M	C	M	C	M	C					
3″	4	0	3	1	6	4	5	0	4	1	5	0	27	6	33	1.83	30.5
5″	5	0	2	2	6	5	5	5	5	0	4	0	27	12	39	2.17	36.1
7″	4	0	5	1	6	5	5	6	6	0	4	1	30	13	43	2.39	39.8
9″	5	0	6	4	8	4	7	5	6	2	5	0	37	15	52	2.89	48.1
ΣM	18		16		26		22		21		18						
ΣC		0		8		18		16		3		1					
Σ(M+C)	18		24		44		38		24		19						
Per cent	25.0		33.3		61.1		52.8		33.3		26.4						

that deviated from it to the extent of .25, .50, and .75 inches, respectively, in both the positive and negative directions.

On critical trials the majority judged contrary to the true relation; for example, it would judge a line 9.75 inches long as shorter than the 9-inch standard. The schema of the comparisons appears in Table 19.

Each critical comparison was made once, giving a total of 24 critical trials. The same comparisons were also shown in the course of the experiment as neutral trials, i.e., the majority judged them correctly. The experiment opened with 8 neutral trials; there were 24 other neutral trials interspersed among the criticals; in all there were 56 trials. Eighteen subjects took part in this variation.

The critical subject had three alternatives: to respond correctly, with the majority, or to compromise. A compromise was in the present context a judgment of "equal."

The relevant data appear in Table 20, which includes the errors in response to each comparison, and lists separately majority (M) and compromise (C) errors.

We note, first, two detailed findings. (*a*) Errors in response to positive and negative discrepancies were almost identical in frequency for each size of standard. Combining the data for all standards the mean number of errors for positive discrepancies was 4.5 and for negative discrepancies 4.8; these were 37.5 per cent and 39.8 per cent of the total number of judgments, respectively. We may therefore in the subsequent analysis combine the data for the two directions. (*b*) Compromise responses varied inversely with size of discrepancy. They increased as the magnitude of contradiction decreased, and became vanishingly low when the discrepancy rose.

Table 21 presents the data of Table 20 in summary form. It gives the errors in per cent according to size of standard and of discrepancy, combining positive and negative directions and treating compromise errors identically with the other errors.

An examination of these data shows the following. (*a*) The frequency of errors varied inversely with the magnitude of discrepancy. A crude index of the relation is obtained by combining the data for all standards; the proportion of errors falls from 56.7 per cent to 33.3 per cent to 25.7 per cent as the discrepancy rises from .25 to .50 to .75 inches. The differences between each of the discrepancy steps are

TABLE 21

FREQUENCY OF ERRORS (IN PERCENTAGES) AS A FUNCTION OF SIZE OF STANDARD AND MAGNITUDE OF DISCREPANCY: EXPERIMENT 6

Magnitude of discrepancy (in inches)	Size of standard (in inches)			
	3	5	7	9
±.25	41.7	58.3	61.1	66.7
±.50	25.0	25.0	33.3	50.0
±.75	25.0	25.0	25.0	27.8

significant. The t value for .25 and .50 inches is 4.01, $p<.01$; for .50 and .75 inches $t=2.64$, $p<.01$. (*b*) Errors increased with size of the standard. The differences were not large but were consistent in direction. The difference between the 3- and 5-inch standards was not significant, nor between the 5- and 7-inch standards. All other differences between the standards were significant, as follows: between 3 and 7 inches, $t=4.63$, $p<.001$; between 3 and 9 inches, $t=4.51$, $p<.001$; between 5 and 9 inches, $t=3.19$, $p<.01$; and between 7 and 9 inches, $t=3.42$, $p<.01$. (Throughout, $df=10$.)

The rise of errors with the size of standard, for which we also obtained evidence in Experiment 1, is a puzzling result for which a rationale is difficult to find. One would suppose that once a difference is clearly above the threshold, variations of relative difference should not affect the certainty of judgment. Our results show, however, that this is the case. They suggest that differences above the threshold do vary psychologically along a continuum of what we might call "unquestionability." This dimension does not emerge in the conventional psychophysical experiment where a small but distinguishable difference seems as certain as a large one. However, as soon as one's judgment is called into question, subjects do take relative differences into account. Their independence then varies systematically within a region far above the psychophysical threshold.

To summarize; (*a*) Errors were a joint function of the magnitude of discrepancy and size of standard. (*b*) Errors rose with size of standard, suggesting that differences above the threshold vary along a dimension of "unquestionability" when judgments are brought under social challenge. (*c*) Errors increased in frequency as the discrepancy of the majority with the given facts decreased.

Experiment 7: The Role of Neutral Trials

The procedure of Experiment 1 provided for the inclusion of neutral trials, to which the majority responded correctly. There were, it seemed to us, a number of reasons for introducing them. By starting with neutral trials we hoped to forestall suspicion, and also accustom the subject to the procedure before he reached the critical trials. We also inserted them in the middle of the series in order to demonstrate that the majority could judge correctly. Nevertheless, our procedure was guided only by an informal sense of what was needed to make the experimental situation credible and effective.

It seemed of interest to explore further the role of neutral trials, both for their psychological significance and for technical reasons. A neutral trial may mark a relief from the tension of disagreement and therefore may heighten independence. Or, yielding to the majority might be the result of continuous application of pressure, which increasingly frequent neutral trials should dissipate. On the other hand, a neutral trial is also an equivocal situation. While it relieves the subject of the odium of dissent (and at times raises his hope that the nightmarish contradiction is coming to an end) it also demonstrates to him that the majority is judging correctly, rendering all the more weird the preceding and following disagreements. If viewed in this way, it might further weaken the subject's resolution.

We accordingly varied systematically the ratio of neutral to critical trials. For each condition there was a different group of subjects, drawn from the three educational institutions described earlier (see p. 12). The following were the variations: (*a*) *Two neutral trials.* These were the two opening trials of the series, identical with the series of Experiment 1. All further neutral trials were omitted, the 12 critical trials succeeding each other without interruption. The ratio of neutral to critical trials was therefore 1:6. One could also think of this condition as eliminating the neutral trials, since they were excluded once the critical trials appeared. There were 14 subjects in this variation. (*b*) *Six neutral trials*, the ratio of neutral to critical trials being 1:2. This was the condition of Experiment 1; here we present the data of that experiment as Experiment 7b. (*c*) *Twelve neutral trials.* The number of neutrals was increased to twelve, which was also the number of critical trials. (The positions of the neutrals among the series of 24 trials was: 1, 2, 5, 7, 8, 10, 13, 14, 17, 19, 20, and 22.) There were 19 subjects in this variation. (*d*) *Twenty-six neutral and six critical trials.* In this final variation we increased the ratio of neutrals to criticals to somewhat over 4:1.[28] Critical trials were separated by four to five neutral trials. (The positions of the critical trials were: 6, 11, 16, 22, 27, and 32.) Because of the length of the series there were only 6 critical trials. There were 14 subjects in this variation.

The results appear in Table 22. While the differences are not significant there is a suggestion that independence increases with a higher ratio of neutrals; comparing the extremes, the errors fall from 53 to 26 per cent. There was also a tendency for compromise errors to increase with the ratio of neutral trials. The proportion of compromise errors on extreme trials was as follows in the several variations: 7a, 5 per cent; 7b, 19 per cent; 7c, 10.5 per cent; and 7d, 41.7 per cent. (These calculations are based on the trials in the first half of the series, to make them comparable with variation 7d.)

We conclude that neutral trials tended to heighten independence, but that their effect was weak. That they did not act more strongly may be, as we suggested earlier, because they initiated both strengthening and weakening forces.

Experiment 8: Varying the Number of Alternatives

The decisive alternatives under the present conditions are between the match which the subject judges to be correct and the incorrect match of the majority. In Experiment 1 a third alternative was introduced. The evidence showed that each of the comparison lengths, including the "silent" one not chosen by the majority, was exerting a definite effect. It is probable that the number (as well as the kind) of alternatives play a part in deciding the outcome. In the following variation we reduced the comparison lengths to two: one identical with the standard, the other differing from the standard and chosen by the majority.

TABLE 22

MEANS OF ERRORS AND PER CENTS: EXPERIMENT 7

Experiment	*N*	Errors	
		Mean	Per cent
7a	14	6.4	53.0
7b	123	4.4	36.8
7c	19	4.6	38.6
7d	14	1.6	26.2

The comparison lengths were those of Experiment 1, except that the line that had not previously been chosen by the majority was now eliminated (see Table 1). For example, Trial 1 retained the 3 and 3¾ inch comparison lines, and omitted the third length. This procedure in no wise changed the estimates of the majority. It reduced the number of alternatives and also abolished the distinction between moderate and extreme errors (in the relational sense they had in Experiment 1).

Immediately following the 18 trials with two alternatives we introduced without pause the first 9 trials of Table 1, with the third comparison restored. (Three subjects did not take part in the second phase of the procedure.) In all other respects the procedure was that of Experiment 1. There were 17 subjects in this experiment.

Table 23 summarizes the distribution of errors. For easier comparison we have

---

[28] We have not studied the most extreme condition, that of excluding neutral trials altogether.

TABLE 23

Distribution of Errors: Experiment 8

Subject	Two comparison lengths (12 trials)	Three comparison lengths (6 trials)
1	0	0
2	0	0
3	0	—
4	0	—
5	1	0
6	1	0
7	1	0
8	2	1
9	2	2
10	2	—
11	5	1
12	10	6
13	12	6
14	12	6
15	12	6
16	12	6
17	12	6

arranged the scores of subjects in the order of increasing errors.

With the alternatives reduced to two—the correct choice and that of the majority—the individual differences ranged from complete independence to complete yielding. Although the level of errors was very similar to that of Experiment 1, accounting for 40.8 per cent of all judgments, the forms of the distributions differ quite strongly. Under the present conditions *the subjects concentrated at the extremes*, intermediate forms of yielding being least frequent. The critical trials, which in Experiment 1 produced characteristically different rates of error, became almost completely homogeneous in this variation. The effect of the procedure was, for reasons not clear to us, to increase the proportion of independent and of strongly yielding reactions.

Table 23 also reports the results for the last trials which included the three alternatives of Experiment 1. Individual performances remained remarkably consistent.[29] With the third comparison line restored, the subjects continued to cluster at the extremes, and the individual differences between the critical trials so prominent in Experiment 1 now completely disappeared.

The consistency of individual performances in the two phases of the variation is best accounted for in terms of an effect of the earlier upon the later condition. We propose that the subjects persisted, despite subsequently altered circumstances, in the direction they first adopted, probably because they felt committed to it. In this manner we can explain the fact that the second phase of this experiment resembles the first and departs strikingly from that of Experiment 1, with which it is identical in procedure. In the light of results to be reported in the future it seems probable that the high consistency of the subjects was in the present case the consequence of a self-commitment.

The variation here reported is only a first attempt at a problem that may deserve systematic exploration. In Experiment 8, two and three alternatives were used; other experiments could use four, five, six, etc. alternatives. Increasing the number of alternatives raises new questions that can be studied in a forthright way, such as their effect on the level of independence and on compromise solutions. Certain figural effects could also be studied in this connection. One might, for example, introduce a gap between the continuous array of alternatives and observe the effect of placing the majority on the extreme or moderate side of the gap. It may also be of interest to vary the position of the majority on the continuum of alternatives while keeping the absolute discrepancy constant.

Experiment 9: The Role of Warning

The conflict with the majority was, under the conditions of Experiment 1, confined mainly to the immediate present. We now undertook to invest the situation with a reference to the future. In brief, we informed the subject that the estimates would be checked by measurement at the end of the experiment. We hoped that this procedure, which we considered to be a counterforce with an objective reference, might throw light on the persistence or fragility of the effect.

---

[29] The reliability of individual consistency on the 2- and 3-line judgments can be demonstrated in the manner described earlier (see footnote 27, p. 58-59). On the assumption of equally-likely random errors the median of the theoretical distribution for 0 to 6 errors is close to 2. For our subjects the difference between error scores for the two conditions is less than 2 for thirteen out of fourteen subjects. (The comparison was between the last 6 trials of the first phase of the variation and the 6 trials of the second phase.) This result differs from chance expectation at p less than .003.

To the instructions of Experiment 1 the following was appended as a last statement: "At the conclusion of the experiment, when you have made the estimates, we will measure the lines with a ruler and show you which of your answers were correct, and which were in error. This we will do as soon as you have completed all the estimates." In addition, the experimenter illustrated the future procedure on the first (neutral) trial. As soon as all had announced their estimates, he measured the standard and the comparison line in front of the group, saying: "Your answers were correct. This is how we will measure the other lengths later." He added that the other comparisons would proceed uninterruptedly and be checked only when all the estimates were in. Thereupon he proceeded to the other comparisons.

In all other respects the procedure was that of Experiment 1. There were 11 subjects in the present variation.

The results resemble closely those of Experiment 1. Errors ranged between 0 and 12; the mean number of errors was 4.4, or 36.7 per cent of all critical estimates. (The individual error scores were: 0, 0, 1, 1, 2, 3, 5, 5, 7, 12, and 12.)

The subjects' comments during the interview as well as our observations agree in indicating that the added instructions had little if any effect in comparison with the impact of the total situation. Completely independent subjects stated that the proposed measurement made them anxious to reach the end, but expressed strong doubt that it affected their reports. A number of yielding subjects expressed the same view, adding that they gave no thought to the matter while the experiment was in progress. No subject stated that the instructions made any difference to his performance.

If one were to take these results at face value it would be in order to conclude that the effect of the majority grows primarily out of the immediately given conditions, that the subject acts mainly in accordance with forces present at the moment. We need however to keep in mind that the warning here employed might not have promoted independence even if a potentially yielding person were constantly aware of it, since the knowledge of a future check would not suffice to relieve him of the doubts that the situation produced. Indeed, we have seen earlier (Experiment 4) that the prospect of a future check was mainly responsible for yielding under circumstances that worked strongly for independence. It would therefore appear that the present procedure maintained intact the conditions of Experiment 1 and that these were not altered appreciably by the addition of a warning lacking in direction. We may conclude that generalized warnings are not effective under our conditions.

We must of course leave open the possibility that other forms of expectation or warning might have definite consequences. We are thinking of procedures such as the following: (*a*) To include in the instructions a specific statement to the effect that majorities are not always right (and, in a more extreme procedure, to add a statement about the hazards of following majority opinion); (*b*) to announce in advance that previous experimentation has shown, contrary to expectation, frequent errors, or to announce that such errors are rare but have significance. In such directions we see the promise of valuable investigation.

Summary

1. The majority effect was obtained with the stimulus dimension of brightness, demonstrating that the effect is not restricted to a particular modality (Experiment 2).

2. Eliminating the public character of the judgment markedly lowered the majority effect (Experiment 4).

3. The high consistency of individual performances found in Experiment 1 was confirmed when the experimental series was tripled in length (Experiment 5).

4. The frequency of errors was demonstrated to vary positively with the size of standard and negatively with magnitude of discrepancy (Experiment 6).

5. Neutral trials had a tendency to

heighten independence, but their effect was weak (Experiment 7).

6. Reducing the alternatives to those represented by the majority position and by the subject's view tended to eliminate intermediate performances, pushing the subjects in the direction of complete independence or yielding (Experiment 8).

7. Generalized warning did not affect the level of independence (Experiment 9).

8. Critical subjects persisted in the direction they adopted early in the experimental session (Experiment 5). They showed a strong tendency to maintain their early direction also when the subsequent conditions were altered in certain respects (Experiments 3 and 8).

VI. STRUCTURE OF THE EXPERIMENTAL CONDITIONS

The crucial feature of the present experiments was the persistent recurrence of a disagreement about a fact between a minority-of-one and a majority. But this was a quite specific, indeed an unusual, kind of disagreement, whose distinguishing characteristics we need to delineate. Having surveyed the results it may be helpful to examine the principal properties of the experimental setting. We may as a result gain a clearer view of the boundary conditions of our data and of their relations to group influences under other conditions. (In this discussion we will be referring mainly to Experiment 1, omitting for the present reference to the variations described in the preceding section.)

1. The circumstance that decisively governed the course of the reactions was the introduction of *an unambiguous objective condition as the object of judgment*. It was this fact that was responsible for the quality of paradox and conflict that the situation engendered. One may nevertheless ask why disagreement about a matter that was in itself not too important aroused the strong reactions that we have described. The answer touches upon the role and function of consensus in social life. Our procedure produced a failure of consensus where it was least understood and expected, tending to turn disagreement into a more ultimate kind of contradiction. Now consensus, especially on fundamental traits of the surroundings, is the vital prerequisite of social action; to abolish or impair it is to threaten the relations of interdependence which ordinarily experience continuously validates. (For a fuller discussion of consensus see 1, Chaps. 4 and 16.) It is in these terms that we propose to account for the strength of the reactions which the present situation produced. Although ostensibly the disagreement centered on very specific and limited data, it acquired a wider import. It signified to the critical subjects that they were at odds with a majority about a basic relation in the world.

2. Another central condition was the *independent access to the facts*. The experimental arrangements insured that each person could see with his own eyes and under optimal conditions. In this respect the present situation differs sharply from other, and frequent, forms of disagreement. Often in social life differences of judgments are about facts that are far less visible. The social and political "lines" and their relations are as a rule not bluntly given in the individual's field. Instead he often depends on others to inform him not only about the interpretation, but also about the existence of facts remote from his experience. This indirectness was here excluded. The facts were constantly present, and the individual could not help but see them as he did.[30]

3. Implied in what has been said is that the minority person accepted the reports of the majority as evidence of their perception and judgment. He knew that they had as direct an avenue to the facts as he did. Further, since he did not (as a rule) question their capabilities or good faith, he could not help but grant their reports the same kind of credence and respect that he wanted to accord and to have accorded to his own observations.[31]

---

[30] On the assumption that the clarity of the perceived data was the source of opposition to the majority, it follows that conditions impairing clear observation should offer less resistance to majority opinion. It should therefore follow that (*a*) successive presentation of standard and comparison lengths and (*b*) brief intervals of exposure would increase the majority effect. Another deduction, that the level of the majority effect would vary with the magnitude of contradiction, we have studied and will report subsequently.

[31] The subject had no ground for questioning the capacities of the majority; they were on the same level as far as the particular task was concerned. The situation would of course alter if doubts on that score became possible, as would be the case if, for example, the majority all wore thick lenses or a patch on one eye, or if their competence were discredited by some other means.

4. *The unanimity of the majority derived added force from the knowledge that it had its source in the independent judgments of individuals.* From the standpoint of the subject the majority was not an interacting group. The type of unanimity is therefore structurally different from that obtained as a result of group decision. Consequently the probabilities of obtaining such unanimity by chance appeared vanishingly low. (To be sure, some subjects came to see the majority as a united opposition, but they were referring not to the source of the contradiction as much as to the presumed reactions of the majority when challenged.)

5. *The experimental situation possessed a relatively closed, self-contained character.* Because the contradiction centered around a specific perceptual relation it was not possible to minimize it by reference to external conditions, such as differences of tradition or attitude. The contradiction was present right within the psychological field.[32]

6. The preceding circumstances shaped the conflict that proceeded from the opposition between "self"-evidence and the unanimous majority. In the first place the conflict had a quality of *irreconcilability.* The facts in question were rigid, not lending themselves readily to restructuring. Nor was there a possibility, for the reasons stated earlier, of an outright repudiation of the majority. Neither was it possible to overcome the disagreement by altering the views of others or by modifying one's own view in order to accord with the group.

7. In addition to its irreconcilability, *the contradiction was ununderstandable, containing a hidden condition which the minority person was unable to penetrate* (provided his suspicion was not aroused). This characteristic is perhaps in greatest contrast to ordinary experience. We almost never meet with opposition, however strange, without possessing some notion about the grounds of disagreement. Consequently we account to some extent for dissent even if we cannot resolve it. Thus we may disagree with others about the merits of a person or a proposal without impugning either their discernment or ours. We may realize that the question is complex, that one or the other side has not taken into account the relevant facts, or failed to weigh them properly. If these explanations do not suffice, we may suppose that interest or bias are responsible. Such considerations were excluded from the present study. Because of its high simplicity the issue was not "controversial." One could hardly dismiss the majority with the convenient cliché that there are two sides to every question. The present situation debarred the minority person not only from reaching a common ground with others; it also prevented him from seeing satisfactorily how the failure of agreement could arise.

8. These conditions created a minority-of-one in a rather special sense. It would be missing the point to say that the present situation differs from usual forms of dissent because it is rare to find in life a minority of one, or that as a rule people hold the most unpopular opinions as members of groups, however small, thus enabling them to face opposition as parts of a social body. However rare it might be, it is not difficult to conceive of a person defending a position single-handedly. More significant is the consideration that even in instances of the latter kind the person is not to himself a minority of one in a final sense. As long as he feels justified in his stand he also has the conviction that if others could be made to understand they would be on his side; at least he can retain faith that he will be justified in the future. Such is the power of our confidence in human reason and feelings (see also 1, p. 359). In contrast, *the loneness with which we are here dealing has a more absolute quality, since the minority of one saw no possibility of persuading or of being persuaded.* The majority, although limited in size, was representative. The minority individual had no reason to suppose that others, not included in the group, would be more likely to side with him. The given majority was symbolic of what any portion of humanity might perceive. These are the reasons why the contradiction struck consternation in many.

This statement needs, however, to be qualified. We have seen that the subjects adopted many measures of defense to mitigate their doubts. Further, the experimental episode arose too suddenly and lasted too briefly to arouse the sinister fear that the incomprehensible disagreement would accompany them through life. The subjects knew that they were in an experiment, and retained a healthy confidence that soon they would be back in the safely familiar world out of which they had been jolted.

9. The situation contained a further significant condition: *the minority of one had to take a stand, to declare himself before the group, when to do so had the meaning of either siding with the group or dissenting from it.* Had the critical subject been only a spectator the unusual discrepancy between the group and himself would most probably have still attracted his attention and given rise to perplexity. But in addition he was required to state his position, a circumstance that intensified his difficulties. (For evidence of the import of the public character of the response, see pp. 56–58).

10. We can now appreciate more clearly the

---

[32] It should, however, be pointed out in qualification that the procedure did limit the disruption of consensus by confining it to a narrow range of data. The contradiction concerned only one detail in the surroundings. Presumably the conditions would have become more drastic if the majority simultaneously challenged the minority-of-one about several independent data in the environment.

forces which the situation produced. There were, on the one hand, the objectively given relations which the subject had undertaken to report accurately, and which therefore exerted a directed demand upon him. Opposed to it was the force of the group. The conflict disrupted irreconcilably and incomprehensibly a consensus about a simple and basic fact in the surroundings. It was continually maintained by the individual's direct experience, which plunged him into the position of a minority of one who could not draw for support upon considerations lying outside the experimental setting, and was further intensified by the necessity of acting publicly. The requirement to state one's position meant that the subject could not escape, postpone, or delegate the difficulty. (No subject at any time refused or excused himself from stating a judgment.) Although the situation was temporary and relatively segregated from the stream of outside events, the minority subject was surrounded by the forces it produced.[33]

11. It is in terms of the relations described above that we need to understand the decision required of the minority of one. He had before him the alternative of adhering openly to his experience and rejecting the majority, or of siding with the majority at the cost of suppressing his direct experience. This is the first meaning of independence and yielding in the present context.

It should be pointed out that in other contexts resistance and compliance can have widely different meanings. Acceding to the will of a group may, under some conditions, require a high measure of steadfastness, and resisting it may be a symptom of weakness. The resistance to group pressure we are here studying is that of a person in a lonely condition opposing an arbitrary group; it is closer to what is known as *courage civile* than to the courage that people show when they act in concert.

There remain to be mentioned the following additional characteristics of the situation.

12. *The minority and majority were peers.* The minority person was only temporarily in a minority in relation to others who were only temporarily in a majority. We have not here studied the problems that would arise if the participants had also differed systematically in other respects—in age, sex, or social position.[34]

13. *The procedure lacked several essentials of a group process.* First, the groups in question were temporary aggregations of persons, gathered for a brief and special purpose. There was consequently, as was noted earlier, a quality of finitude to the situation which contrasts significantly with the chronic and often irrevocable relations of social life. Second, the present groups lacked the property of continuous mutual interaction. Since the majority performed a predetermined function it did not form, together with the minority of one, a self-regulating unit. There was social action, but it was confined mainly to the acts of the critical subject.

14. Another aspect of the preceding point may now be stated. *The group did not exert pressure in the usual sense of persuading or applying sanctions.* There were no rewards or punishments in the ordinary sense of these terms, nor were there any specific "interests" introduced which might serve as goals or as a basis for calculation.[35] Rather the subject decided for himself what his interests were, and what mode of action he found rewarding or punishing. Objectively the majority was an entirely neutral chorus of voices. Whatever pressure the subject felt grew solely within himself.[36]

15. *The task concerned a physical fact in the surroundings which was not intrinsically related to the constitution of the group.* Doubtless the dynamics would be different in many respects if the focus of contradiction were itself social and related to the group structure.[37]

VII. Summary and Conclusions

1. We investigated some of the conditions responsible for independence and lack of independence in the face of arbitrary group pressure. To this end we produced a disagreement between a group and one individual member about a clear

---

[33] It should be understood that the forces in question were within the phenomenal field of the subject and not in the surroundings.

[34] One subject in a related experiment, who was independent, remarked in the course of the interview with a superior air that he was the only engineering student in the group. In general, the course of events might be considerably modified once it became possible to adopt an evaluative attitude toward the majority.

[35] To introduce experimentally specific incentives or interests—either individual or social—for siding with the majority (or for opposing it) might be of value, especially since we could compare the effects with those obtained under the present conditions.

[36] The situation was also most highly personalized in the sense that the subject acted solely for himself. He had neither the responsibility nor the opportunity to act on behalf of others, a circumstance that might have exerted marked effects.

[37] The issue is not solely or even mainly whether the question in dispute concerns a physical or social datum. There are social data at least as unambiguous as those in the physical sphere; for example, the meaning of a word or phrase, or the relative dates of the Roosevelt and Truman administrations. Far more significant is the relation of the given issue to the goals and structure of the group.

and simple issue of fact. The majority, which was in all cases unanimous, co-operated with the experimenter by reporting wrong judgments, in disagreement with the individual member (or the critical subject) who could only judge the facts correctly. The critical subject was called upon to state his judgments in public, after hearing the report of the majority.

2. The task required the comparison of the relative length of lines. A standard line was to be matched with one of three comparison lines which differed appreciably in length from one another, and of which one was equal to the standard. The unanimous majority, which consisted of 7 to 9 persons, as a rule matched the standard with a length that differed from it. The errors of the majority were far above the threshold.

3. The contradictions by the majority deflected considerably the estimates of the minority in its direction. Whereas the judgments were virtually free of error under control conditions, one-third of the minority estimates were distorted toward the majority. At the same time the minority estimates remained preponderantly accurate, testifying to the force of the perceived relations under the given conditions (Experiment 1).

4. Individuals differed markedly in resistance to the arbitrary judgments of the majority. The performances ranged from complete independence to complete acquiescence.

5. The majority effect was not restricted to the relation of length. It was also obtained when the contradiction concerned brightness relations (Experiment 2).

6. The pressure of the majority did not increase with time. Most subjects remained at a constant level of independence throughout the experimental episode (Experiments 1 and 5).

7. The majority effect here obtained was a function of the public character of the proceedings. When the minority subject was relieved of the necessity of announcing his dissent openly the level of errors dropped markedly (Experiments 1 and 4).

8. There was a far-reaching lawfulness in the operation of the majority effect. The majority effect was shown to be a function of (*a*) the stimulus-conditions, (*b*) of the majority position, and (*c*) of factors residing within the subjects.

9. The stimulus-conditions controlled the operation of the majority effect in a variety of ways. (*a*) Comparable experimental groups responded similarly (Experiment 1). (*b*) Identical stimulus constellations produced consistently similar effects (Experiment 1). (*c*) The occurrence of certain errors (such as compromise reactions) was a function of the particular stimulus relations (Experiment 1). (*d*) The level of errors decreased as the magnitude of the majority contradiction increased (Experiment 6). (*e*) Errors increased in frequency with increase in the absolute size of the standard (Experiments 1 and 6).

10. The action of the majority controlled the level and quality of errors. A moderately erring majority induced only moderate errors, while a majority making extreme errors induced an appreciable number of compromise errors (Experiment 1).

11. The performances of individuals were highly consistent. (*a*) Individuals showed a marked tendency to be consistently independent, yielding, or intermediate in coping with the pressure of the majority. The performances during the early and late phases of the experimental episode were significantly associated (Experiments 1 and 5). (*b*) There was a high association between the latency of submission to the majority and subsequent independence (Experiments 1 and 5). (*c*) Individual performances remained

highly consistent in the face of certain changes in the given conditions. When subjects established a level of independence they tended strongly to maintain it when the circumstances altered within the limits here employed (Experiments 3 and 8).

12. The interview, which followed the experimental session, provided qualitative evidence concerning the effects produced by the majority. (*a*) Subjects endeavored to resolve the conflict, of which all were aware, by means of various hypotheses which they were compelled to surrender as the contradiction continued. (*b*) The contradiction frequently produced concern, doubt of one's accuracy, and temptation to join the majority. (*c*) As the opposition persisted the reactions became increasingly self-centered. Subjects expressed fear of conspicuousness, of public exposure of personal defects, and of group disapproval; they felt the loneliness of their situation.

13. The subjective reactions to the conflict of independent and yielding subjects differed markedly in certain respects. (*a*) Conviction of rightness, freedom from doubt, and absence of temptation to join the majority were more frequent among independent subjects. The differences, although highly significant, did not, however, completely separate the independent and compliant groups. There was no relation between feelings of concern and independence. (*b*) Yielding subjects seriously underestimated their compliance. They offered a variety of reasons for their errors, the most usual being the painfulness of standing alone against the majority.

14. Subjects offered diverse grounds both for independence and yielding. On the basis of the comments of those whose performances were extreme a number of outstanding modes of reaction were described. (*a*) The independence of some appeared due to a basic confidence in the face of group opposition. (*b*) There were, however, subjects who, while convinced that the majority was right, remained independent. (*c*) A very few yielding subjects appeared unaware of the effect of the majority upon them. (*d*) A substantial proportion of subjects yielded once their confidence was shaken. The presumed rightness of the majority deprived them of the resolution to report their own observations. (*e*) Others who yielded lost sight of the question of accuracy, being dominated by an imperious desire not to appear different, apparently out of fear of revealing a general and undefined defect.

15. The problem of individual differences in independence and their possible relations to character and early experiences was discussed.

16. The particular properties of the experimental situation and their relation to more usual social contradictions were described.

REFERENCES

1. Asch, S. E. *Social psychology.* New York: Prentice-Hall, 1952.
2. Asch, S. E. Effects of group pressure upon the modification and distortion of judgment. In H. Guetzkow (Ed.), *Groups, leadership and men.* Pittsburgh: Carnegie, 1951. (Also in G. E. Swanson, T. M. Newcomb, & E. L. Hartley (Eds.), *Readings in social psychology.* New York: Holt, 1952).
3. Mosteller, F. & Bush, R. R. Selected quantitative techniques. In G. Lindzey (Ed.) *Handbook of social psychology*, Cambridge: Addison-Wesley, 1954. Chap. 8.

(*Accepted for publication October 19, 1955*)

Vol. 70, No. 10 Whole No. 417, 1956

Psychological Monographs: General and Applied

Fourteen Dimensions of Temperament

J. P. Guilford and Wayne S. Zimmerman
University of Southern California *San Diego State College*

The Problem

One of the most widely used approaches to the assessment of personality traits in individuals has been that of the personality inventory. In this approach, the overt indices of traits, the items of information concerning the person from which trait values are inferred, are in the form of his responses to questions or statements about his habits, his likes and dislikes, and his opinions. In interpreting this information, although much attention is given to the literal meanings of the questions and answers, the psychological traits indicated are more properly decided on the basis of empirical evidence. The empirical evidence has often been based upon correlations among the items, and between the items and selected criteria. Sometimes the summarizing device that is applied to the correlations has been that of factor analysis.

The earlier factor analyses by the senior author and others (5, 6, 7, 10, 12) were based upon the intercorrelations of single items. When there was little known as to which item would indicate a trait in common, this was the best procedure. Also, at that time, the number of inventory items in existence was much smaller than now.[1] After determining the underlying factors and their apparent properties, the senior author proceeded to develop three personality inventories: *An Inventory of Factors STDCR*, *An Inventory of Factors GAMIN*, and *The Personnel Inventory*, with H. G. Martin as co-author of the last two (2, 8, 9). In their development, new items were written for each factor. These, and the old ones, were put through the process of internal-consistency item analysis to form pools of items, with scoring weights, to develop a measuring scale for each factor, primary trait, or primary dimension of personality.

Before going further, it is desirable for us to review the names of the factors. The factors of *Inventory GAMIN* will be mentioned first, in agreement with the order used hereafter in this report:

G —General activity
A —Ascendance vs. submission
M —Masculinity vs. femininity
I —Confidence vs. inferiority feelings
N —Calmness, composure vs. nervousness
S —Social introversion-extraversion; seclusiveness vs. sociability
T —Thinking introversion-extraversion; reflectiveness
D —Depression vs. cheerfulness
C —Cycloid disposition vs. emotional stability
R —Rhathymia; unconcern vs. seriousness
O —Objectivity vs. subjectivity
Ag—Agreeableness vs. belligerence
Co—Cooperativeness, tolerance vs. criticalness, fault-finding

The present report is concerned with a comprehensive factor analysis that involves all 13 factors. A new analysis seemed very much needed for several reasons. The analyses on which the inventory construction had been based were limited in scope. In no one analysis, to our knowledge, have

---

[1] Rumor has it that there now exists a file of some 10,000 inventory items at one of our major universities.

Bureau Ednl. Psy. Research

all 13 factors been properly represented. The senior author's analysis on which the three factors of the *Personnel Inventory* were based was never published.[2] In recent years some questions have been raised concerning whether the same area of temperament can be accounted for on the basis of a smaller number of factors (11, 13). The existence of some of the factors has been questioned. For example, in listing what he considers to be the established factors, French (1) does not include the factor Co.

The evidence that might be interpreted to indicate a smaller number of factors than 13 will now be reviewed very briefly. Lovell (11) intercorrelated the 13 inventory scores and extracted six factors, using estimated communalities in the diagonal cells of the correlation matrix. She appropriately interpreted the resulting dimensions as being "super-factors."

To some extent, Lovell's super-factors account for actual intercorrelations among the 13 psychological factors. In this sense, they would approach what Thurstone calls "second-order" factors. In our opinion, however, they are by no means identical with the second-order factors that one would find by an ordinary oblique-rotation solution involving the 13 first-order factors. The correlations among the 13 inventory *scores* are undoubtedly higher than the correlations among the 13 *factors*. In part, the additional covariances among scores arise from the fact that the score variables are not experimentally independent. In *Inventory STDCR*, and to some extent in *Inventory GAMIN*, and as between the two, many items are weighted similarly in two or more keys, since many items themselves are indicators of two or more factors. Thus, some specific and error variances represented in certain items contribute spuriously to intercorrelations. Another source, even when items are keyed for only one factor, is the tendency for items scored for one factor to be loaded systematically on some other factor. Lovell was aware of these facts and appropriately concluded that the super-factors represent "factors" common to the inventory *scores*.

In his analysis, Thurstone asked the appropriate question as to the actual dimensionality represented by the 13 scores. Using Lovell's correlation matrix, he entered estimates of reliability in the diagonal cells. This introduces specific variances into the picture, making it possible, under the right conditions, for 13 factors to appear.

Unfortunately, in addition to the already-noted lack of experimental independence of the 13 scores and the factorial complexity of items, and their effect upon the intercorrelations, there is a further difficulty stemming from the fact that Thurstone used estimates of reliability presented in the test manuals; in other words, not those based upon the same sample from which the intercorrelations were obtained. In this type of solution, the number of factors obtained depends very much upon the accuracy of the estimates of reliability and upon their consistency with the intercorrelations. It is unlikely that the reliability estimates he used are inflated in keeping with the intercorrelations. As a matter of fact, they are more likely to be underestimates. If they are too small the number of factors is also likely to be underestimated. In fairness to Thurstone, it should be said that he discusses these sources of error, but he concluded that they would not invalidate his analysis (13, p. 13).[3]

Thurstone found nine dimensions in the factor structure of his solution. Seven of these were given psychological interpretations, with names that are somewhat different from those used for roughly corresponding factors listed above. The remaining two he regarded as residual factors. Thus, the seven accepted factors fall far short of the hypothesized 13.

PROCEDURE

In attempting to determine whether a certain factor exists, it is good practice in a factor analysis to make sure that there are at least three experimental variables present to represent that supposed variable. We therefore applied this principle to the study of our 13 hypothesized factors.[4] Furthermore, for a more convincing test of the hypotheses, it was decided to base the analysis upon the same fundamental data as Lovell and Thurstone had used. Dr. Lovell placed at our disposal the original test booklets, which fortunately had been preserved.[5]

---

[2] It can be reported here that this analysis had been initiated in collaboration with Dr. Don W. Dysinger, as a consequence of the clinical insight of Dr. Roswell H. Johnson. In his use of the *Humm-Wadsworth Temperament Scale* in counseling, Dr. Johnson became convinced that the Paranoid score involves three distinct variables. This conclusion was verified by the factor analysis. In a published report, Mosier (12) has reported Factor O, interpreting it in terms of its opposite pole, hypersensitivity.

[3] The senior author has expressed elsewhere at length his views concerning the use of total scores from inventories in general in factor analyses (3).

[4] The initial steps in this investigation were a joint effort of the senior author and Earl L. Taylor, whose fatal illness prevented his continuing beyond the stage of factor extractions. We should have been glad to list him as co-author, but felt that it would be unfair to commit him to expressions of views of which he might not have approved.

In obtaining several scores for each factor, we might have divided the keyed items for each factor randomly into three or more pools. We rejected this idea in favor of a more rational approach; an approach that would give us more information concerning the nature of the factors. By inspection of content and from consideration of relevant, previous statistical information, we combined items into small, homogeneous groups. Two investigators[6] independently categorized the items for each factor, then, after discussion, eliminated any items on which there was disagreement. The resulting score variables are listed below.

The score variables. Each score variable, or test, is indicated by the factor letter symbol with a subscript followed by a descriptive title. A sample item is given for each variable.[7] The number in parentheses indicates the number of items in the test.

1. G_1 *Rapid Pace:* Do you usually eat more rapidly than the average person, even though there is plenty of time? (6)
2. G_2 *Drive for Activity:* Are you the kind of person who is "on the go" all the time he is awake? (3)
3. G_3 *Energy:* Do you often feel bubbling over with excess energy? (2)
4. G_4 *Liking for Action:* Are you happiest when you get involved in some project that calls for rapid action? (3)
5. G_5 *Quickness:* Are you inclined to be quick in your actions? (2)
6. A_1 *Being Conspicuous:* Do you dislike to have people watch you while you are working? (4)
7. A_2 *Maintaining One's Rights:* Do you ever protest to a waiter or clerk when you think you have been overcharged? (10)
8. A_3 *Self Defense:* Are you rather good at bluffing when you find yourself in difficulty? (4)
9. A_4 *Social Initiative:* Do you ever take the initiative to enliven a dull party? (6)
10. A_5 *Fear of Social Contacts:* Have you ever been hesitant to make applications for a job in person? (6)
11. M_1 *Fearfulness:* Are you afraid of snakes? (11)
12. M_2 *Inhibition of Emotional Expression:* Do you cry rather easily? (7)
13. M_3 *Masculine Vocational Interests:* Would you rather be a florist than a miner? (5)
14. M_4 *Masculine Avocational Interests:* Do you prefer going to a dance rather than to a prizefight? (8)
15. M_5 *Disgustfulness:* Do odors of perspiration disgust you? (6)
16. M_6 *Sympathy:* Do you feel deeply sorry for a bird with a broken wing? (3)
17. I_1 *Personal Strengths and Weaknesses:* Are you physically stronger than most of your friends of the same sex? (7)
18. I_2 *Feeling of Adequacy:* Do you always know what to do next? (7)
19. I_3 *Self Confidence vs. Inferiority Feelings:* Do you always feel that you can accomplish the things you want to do? (6)
20. I_4 *Discontent with Self and Status:* Do you ever wish that you were taller or shorter than your are? (9)
21. I_5 *Feeling of Acceptance by Others:* Are you frequently afraid that other people will not like you? (7)
22. N_1 *Tense-excited vs. Calm-relaxed:* Can you relax easily when lying or sitting down? (6)
23. N_2 *Restlessness:* Can you usually sit still without fidgeting? (6)
24. N_3 *Nervousness and Jumpiness:* Do long-continued noises "get on your nerves"? (8)
25. N_4 *Annoyance and Irritability:* Did you ever become so angry at your work that you threw tools about or broke them? (8)
26. N_5 *Fatigueability:* Do you wake up feeling tired in the morning? (6)
27. S_1 *Liking for Friends and Acquaintances:* Do you enjoy getting acquainted with most people? (4)
28. S_2 *Social Leadership:* In social conversations, are you usually the listener rather than the talker? (4)
29. S_3 *Social Poise:* Does it embarrass you a great deal to say or do the wrong thing in a social group? (5)
30. S_4 *Liking the Limelight:* Do you generally feel uncomfortable when you are the center of attention on a social occasion? (4)
31. S_5 *Shyness, Bashfulness:* Do you have to fight against bashfulness? (3)
32. S_6 *Gregariousness:* Do you like to mix socially with people? (4)
33. S_7 *Liking for Social Affairs:* Do you like to have many social engagements? (2)
34. T_1 *Meditativeness:* Do you like to have time to be alone with your thoughts? (4)
35. T_2 *Liking for Serious Thinking:* Do you like to

---

[5] We wish to express our thanks to Dr. Lovell for her generosity.

[6] Mr. Taylor and the senior author.

[7] Items are quoted by permission of the copyright holder.

discuss the more serious questions of life with your friends? (4)

36. T_3 *Analysis of Self and Others:* Do you often speculate about why people behave as they do? (4)
37. D_1 *Emotional Depression:* When you stop to consider your future does it seem very optimistic? (4)
38. D_2 *Worry, Anxiety:* Have you ever lost sleep over your worries? (2)
39. D_3 *Cheerfulness:* Are you usually in good spirits? (2)
40. D_4 *Loneliness:* Do you feel lonesome even when with other people? (2)
41. D_5 *Physical Depletion:* Have you often felt listless and tired for no good reason? (2)
42. D_6 *Feelings of Guilt:* Do you often have a feeling of unworthiness? (2)
43. C_1 *Emotional Fluctuations:* Do you ever change from happiness to sadness or vice versa, without good reason? (5)
44. C_2 *Emotional Excitability:* Are you often in a state of excitement? (4)
45. C_3 *Emotional Perseveration:* Do you find it difficult to go to sleep at night because experiences of the day "keep running through your head"? (4)
46. C_4 *Emotional Immaturity:* Does it upset you much to lose in a competitive game? (6)
47. C_5 *Absentmindedness:* Does your mind often wander while you are trying to concentrate? (3)
48. C_6 *Daydreaming:* Do you often think or dream of what you will be doing five years from now? (8)
49. R_1 *Carefreeness vs. Restraint:* Are you ordinarily a carefree individual? (4)
50. R_2 *Impulsiveness:* Are you inclined to act on the spur of the moment without thinking things over? (5)
51. R_3 *Seriousness vs. Unconcern:* Do you ever take your work as if it were a matter of life and death? (3)
52. R_4 *Liking for Action and Excitement:* Do you often crave excitement? (3)
53. R_5 *Reticence:* Do you dislike to talk about yourself, even to close friends? (2)
54. R_6 *Rapport with the Environment:* Do you become so absorbed in watching an athletic contest that you completely forget yourself? (2)
55. O_1 *Egocentrism:* Do you often feel that a lecturer is talking about you personally? (8)
56. O_2 *Hypersensitivity:* When you are criticized does it disturb you badly? (5)
57. O_3 *Ideas of Reference:* Are you bothered by the idea that someone is reading your thoughts? (5)
58. O_4 *Unwarranted Sympathy:* Do you sometimes feel sorry for all the people in the world? (4)
59. Ag_1 *Contempt for Others:* Do you sometimes think that most people are stupid? (8)
60. Ag_2 *Resistance to Control:* Have you very much resented having members of your family or friends give you orders? (4)
61. Ag_3 *Hostility:* Is it difficult for you not to bear a grudge against someone who has injured you? (7)
62. Ag_4 *Overt Aggression:* Do you enjoy taking part in a good fight? (5)
63. Co_1 *Faultfinding with Human Nature:* Do you believe that most people shirk their duties whenever they can do so? (11)
64. Co_2 *Faultfinding with Society:* Do you believe that only people with money can get a square deal in the courts of law? (6)
65. Co_3 *Faultfinding with Industry:* Do many men deserve higher pay than their bosses? (6)
66. Co_4 *Trusting Honesty in Others:* Can most people be trusted completely? (5)
67. Co_5 *Suspicion of Hypocrisy:* Do you think that most people who help others secretly dislike doing so? (6)
68. Co_6 *Suspicions of Actions:* Were you ever ignored or "given a raw deal" through spite? (7)
69. Co_7 *Victim of Hard Luck:* Does it seem to you that other people "have all the luck"? (5)
70. M_7 *Male Sex Membership.*

Testing and scoring. The sample of students tested by Lovell included 126 men and 87 women, or a total of 213. They were tested under what would seem to be favorable conditions for obtaining truthful cooperation. Since both sexes were represented, we added to the experimental variables the male-female dichotomy, which was correlated with all score variables by estimating point-biserial coefficients. The inclusion of the sex-membership variable was designed to pinpoint the masculinity-femininity factor, thus to clarify the factor structure in other respects.

The items in each test were scored consistent with the key for its factor, with uniform scoring weights of one and zero. A high score for each test was thus indicative in the direction of the trait as named previously in the list of the 13 factors. This meant that all variables were scored in the more socially approved direction except in the case of factors S, T, D, and C (assuming the extroverted and emotionally adjusted poles in each case to be more socially approved).

The correlation matrix. Since most of the score variables were derived from small numbers of items, sometimes as small as two, we dichotomized every score distribution. Assuming that the distribution of the quality represented by each score variable is normal in the population, we estimated tetrachoric coefficients among the scores, by means of the cosine-pi approximation. Since the points of dichotomy were generally in the region of the medians, the approximations may be accepted as having negligible biases.[8]

The factor analysis. Eighteen centroid factors were extracted using punched-card equipment.[9] Extractions were continued so as to include a liberal number of factors. A few residuals as large as .10 remained, but with a sample of about 200 and with tetrachoric r's, it was thought that little or no genuine common-factor variances were overlooked. As it turned out, after rotations four factors were regarded as residuals. The centroid factor matrix is given in Table 1.

Axes were rotated by the Zimmerman graphic, orthogonal method (14), the major criterion being simple structure. Since the reflection of the S, T, D, and C variables led to a correlation matrix with few negative coefficients, the criterion of positive manifold was also a guide. Although there is probably some correlation among some of the factors (to be noted later), we preferred orthogonal rotations as a better approach to a verification study of the factors. It has been found by experience that orthogonal rotations yield more reproducible results (results less dependent upon the composition of the particular test battery). We are also of the opinion that the correlations of oblique axes in a particular analysis are not necessarily good estimates of correlations between psychological factors.

The rotated factor matrix is given in Table 2. It will be noted that a very liberal number of zero loadings was achieved, indicating good, orthogonal, simple structure.[10] Fourteen of the factors may not be regarded as residuals, hence the dimensionality represented in this analysis is at least 14.

INTERPRETATION OF THE FACTORS

All 13 of the hypothesized factors were essentially verified. The definition and scope of two factors—C and R—are somewhat different, however, and a fourteenth factor emerged, borrowing some properties formerly attributed to Factors C and R. In general, the use of meaningful, homogeneous scoring variables helps very much to clarify the meanings of most of the factors.

In discussing each factor, in turn, we first list the tests that have significant loadings on the factor. A significant loading for interpretation purposes is arbitrarily defined as being .30 or greater, positive or negative. For each test variable is given its number, its name, its loading on the factor under discussion, and finally significant loadings for that test on other factors. The algebraic sign of a loading is stated with respect to the factor name and the test trait. That is, where a loading is positive, there is a positive correlation between the quality described in the trait name and the quality of the factor.

Where another factor tends to be repeatedly loaded in tests of a certain factor list, the loadings are listed in a column to

---

[8] The correlation matrix is not included in this report, to conserve on publication space.

[9] We are indebted to Mr. Norman W. Kettner for supervising this part of the investigation.

[10] The algebraic signs for score variables S, T, D, and C are consistent with the general reflections of those variables in the correlation matrix.

TABLE 1

CENTROID FACTOR MATRIX*

Test	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	h^2
1	−03	−05	31	−01	−23	18	−19	−24	19	−16	25	16	17	11	−03	−11	07	−03	49
2	−04	06	44	−07	−19	17	−14	−22	−10	12	−14	−09	17	07	18	−05	10	20	51
3	18	09	35	08	20	10	−26	−26	21	−16	12	−19	−12	10	08	−08	12	07	53
4	02	−20	35	21	−09	12	−11	−19	25	22	15	07	−11	−08	12	16	−13	09	50
5	18	22	24	−08	13	−19	−27	−10	−14	09	19	13	−20	06	05	02	−11	−15	44
6	50	09	38	−12	−21	−11	10	10	−14	−11	−13	21	−10	07	06	22	03	−04	66
7	30	−19	37	−15	−12	−26	−10	29	−19	25	18	−33	14	19	13	29	−10	−04	87
8	40	−20	29	−19	−29	−13	−22	14	−08	−07	−08	−09	14	20	−11	−16	−09	−02	62
9	38	12	57	−35	03	17	−07	08	−08	11	18	17	−10	−13	12	04	−10	05	78
10	41	15	43	05	−14	−08	14	−08	−11	−08	−11	08	06	11	−21	10	−11	02	55
11	43	−40	05	10	−17	10	−16	08	−29	−23	−10	30	−06	−04	−05	−05	−09	−06	69
12	39	−41	−11	−12	−02	18	−12	−21	−01	−05	−11	11	−15	−17	12	−08	−09	05	55
13	32	−58	−09	20	−21	24	07	−11	−19	10	12	−15	−14	−11	06	05	−08	−16	76
14	27	−65	−16	19	06	09	−19	−02	−26	−09	06	−04	05	−12	02	14	−04	−20	76
15	07	−19	−13	29	−12	08	−22	27	−26	−20	−01	08	−22	−12	11	−11	−23	05	54
16	29	−46	02	−04	−10	−07	23	15	13	11	18	24	21	−09	15	−26	12	−12	68
17	44	−09	30	−03	−16	13	−07	−15	43	−11	13	−12	07	02	−10	−05	−06	−20	65
18	58	−11	31	−16	−27	−02	−13	−16	−09	−07	−04	11	−02	−12	03	−01	20	−12	68
19	51	−14	32	−28	−16	−09	06	−24	15	−15	−17	17	−05	−02	−02	06	13	10	69
20	68	13	14	−12	−21	06	02	−16	10	−13	−20	−20	09	−07	18	−05	02	−07	75
21	55	27	10	−07	−27	−06	25	−09	11	−18	11	−18	−13	03	−06	27	−06	−07	73
22	56	−27	−11	09	16	−10	22	13	08	−32	06	−19	−10	09	−11	03	09	25	76
23	57	01	−07	−15	11	−29	15	13	−03	−01	19	04	−12	13	−20	−18	09	−06	64
24	68	−11	−23	−15	08	07	−09	16	03	−22	04	−16	−13	30	17	−11	03	22	87
25	55	14	−30	05	−08	30	−12	09	18	−20	−11	−06	−15	10	−09	−16	06	09	70
26	49	09	−13	−17	23	−04	−19	10	21	−12	17	18	04	18	19	−08	−13	−20	65
27	46	31	22	09	08	14	24	24	12	17	−11	−15	−04	−17	−10	18	−15	−20	72
28	38	29	50	−19	−02	11	−12	28	−17	13	20	−07	−10	−04	−17	−07	−08	12	78
29	62	10	37	−17	−09	−10	08	−09	−10	−11	15	−11	−09	−06	−15	−16	−05	16	74
30	46	38	42	−17	14	−12	−16	20	−10	21	−12	01	−10	−17	−10	12	15	16	84
31	55	38	35	−11	−13	−21	−03	13	05	06	−20	−12	−06	−13	−15	−05	−15	05	79
32	17	24	43	12	08	22	32	−10	08	18	−09	11	09	10	12	07	14	17	60
33	30	18	52	29	22	24	16	−10	13	18	05	14	−13	−10	−19	−26	−09	22	88
34	56	−11	−12	−12	24	04	−16	−11	10	23	−13	07	−06	21	−06	13	07	05	61
35	31	03	−23	−10	41	−02	−16	−33	−19	11	−28	−06	−09	29	−04	−16	−18	−10	75
36	13	−05	−11	08	14	−17	10	−12	−09	16	−30	05	15	14	05	14	−11	−07	32
37	50	−06	04	−12	21	26	−07	23	−03	−18	−07	20	15	18	−11	05	32	10	70
38	58	−26	−18	06	37	−29	14	−02	02	−08	14	12	09	−08	−18	−14	03	−04	79
39	44	04	10	−15	43	25	23	14	−09	−22	−19	−11	11	−07	16	−10	16	−21	77
40	62	02	−19	−15	09	14	−07	−13	06	−02	−29	02	20	−07	−07	10	−24	09	71
41	67	03	−12	−16	28	13	−06	−09	05	−07	12	13	10	−12	−04	09	08	−10	69
42	61	16	−09	−23	11	−14	13	−15	−16	−13	18	03	13	−16	08	05	−17	11	70
43	61	−10	−05	−15	31	36	03	10	21	05	10	−06	14	−14	−04	17	11	08	79
44	47	−38	−28	−12	−26	20	22	−06	−10	13	−08	09	−23	−08	−07	−05	06	−09	74
45	49	−24	−28	−27	16	−09	08	06	09	10	15	18	−03	−21	−29	08	14	−13	74
46	25	05	−27	−27	−26	20	33	−10	−19	−15	07	−16	17	−06	−06	−03	11	−13	59
47	57	−05	−13	−43	17	−03	11	−15	16	24	13	01	−08	06	23	−19	05	14	81
48	56	−09	−17	−28	−12	−05	−14	−22	18	13	−14	06	04	09	−20	07	−08	−07	65
49	15	−09	41	49	22	−10	07	−19	−17	−09	−15	−17	−10	06	09	−13	13	−11	69
50	15	−12	−43	−21	−15	23	38	08	−07	−10	−13	−06	17	−12	12	−01	−14	30	69
51	22	−27	20	39	23	−31	02	06	09	03	−10	−24	−05	08	10	−08	−10	−16	60
52	02	−18	37	32	32	10	−01	−12	−04	−13	10	−16	11	07	−12	−08	−13	14	53
53	31	−05	29	14	−10	−10	12	13	15	13	20	23	13	14	07	10	−10	04	45
54	23	13	28	24	20	−07	−24	−18	−06	15	01	02	18	−19	13	−17	03	12	50
55	58	14	−16	12	−11	−19	14	−16	−12	−21	21	−06	−06	16	13	10	−04	09	66
56	65	−08	−06	−09	−15	−19	02	10	02	07	−13	−27	06	−07	09	−06	−13	20	68
57	57	12	−21	06	−12	−26	−16	−16	−06	−34	15	14	−13	−11	10	10	08	08	74
58	37	−34	−16	06	−14	−21	−12	−23	−05	14	−04	−16	14	−18	−18	−03	18	05	58
59	26	39	−38	32	−08	−04	06	31	03	08	−22	14	11	−04	11	−04	06	−05	68
60	53	39	−13	17	−07	11	−13	13	23	−05	−10	−11	−12	−16	−07	19	08	−16	72
61	42	32	−25	07	−17	−21	−09	12	34	−20	−16	16	−03	06	10	02	−03	08	67
62	41	29	−29	06	−12	−21	−04	−09	31	10	−03	−15	07	−10	10	−13	23	−02	63
63	53	25	−37	27	−12	25	−08	06	−11	19	20	−18	04	09	−18	02	13	−04	82
64	44	03	−16	17	−07	05	05	−21	−04	11	−03	06	18	20	−13	−19	−16	10	48
65	29	29	−43	18	−18	13	−11	12	−05	18	01	12	43	07	−14	−04	−02	−06	73
66	39	25	−32	27	−19	23	16	−09	−05	13	07	19	−19	23	08	15	22	−07	74
67	55	40	−31	22	−24	06	09	06	11	07	20	06	−04	13	10	−12	05	−06	79
68	49	36	−39	02	07	−06	−30	10	−07	−14	05	−03	18	−20	14	13	07	14	79
69	45	20	−28	−08	−13	−34	−16	11	−08	21	02	−10	−13	06	08	−08	09	−07	60
70	27	−63	−09	23	−21	15	−05	08	−18	17	−09	03	−13	−07	11	−04	08	−08	72

---

* Decimal points have been omitted.

TABLE 2

ROTATED FACTOR MATRIX*

Test	G	A	M	I	N	S	T	D	C	R	O	Ag	Co	C_1	R_1	R_2	R_3	R_4
1	47	06	10	23	−04	−07	−27	08	−14	−03	−01	−24	08	−06	10	09	−09	08
2	42	26	−01	00	−10	08	01	−21	−16	19	02	−21	00	18	18	−16	09	06
3	44	−05	−17	17	24	09	−05	03	−14	36	12	−12	02	00	−17	05	−09	02
4	37	04	20	07	−19	13	02	−12	−17	08	07	−02	−18	05	−20	34	−09	05
5	18	21	−10	12	16	18	18	00	−11	−08	10	−13	13	−22	−26	−01	−08	−27
6	07	49	08	41	09	26	06	17	−07	−06	07	16	−08	04	−08	−23	17	−03
7	−07	74	16	09	25	−05	09	−18	06	24	06	−07	−04	−01	−05	22	18	−19
8	16	46	25	22	37	02	02	01	13	04	17	−06	−05	−09	29	−08	−03	05
9	29	48	−02	16	14	48	−05	12	−01	−06	−03	−17	−08	21	−14	01	−06	−24
10	07	40	−01	43	−05	29	14	15	−01	12	08	−01	01	−10	06	−11	01	18
11	03	17	60	13	15	09	01	30	01	−08	18	00	−08	−02	−05	−29	−13	06
12	20	−10	39	08	12	02	13	17	16	−05	25	01	−15	32	−08	−11	−17	−07
13	−01	−03	72	14	−06	−03	04	−05	28	20	07	−01	−05	15	−21	05	−03	−05
14	−08	04	64	−08	04	−26	06	26	16	22	18	−06	−08	00	−23	−04	−09	−09
15	−21	−03	45	−13	17	12	−13	−08	−15	−02	12	11	−03	−08	−12	−21	−28	02
16	01	−01	48	14	06	08	−10	28	02	−08	12	−01	−08	13	30	30	25	−26
17	44	13	09	44	11	05	−12	17	20	21	06	10	−05	−02	08	25	−17	−01
18	38	31	25	34	09	15	−04	20	17	01	30	02	−01	06	−02	−23	14	−14
19	34	23	03	45	10	11	07	26	08	−07	30	02	−23	21	05	−11	15	06
20	28	22	02	43	12	12	03	09	19	26	23	29	09	27	16	−20	−05	−14
21	−01	27	−12	64	02	12	−07	01	23	10	10	27	13	11	−17	00	00	03
22	−28	−03	13	30	40	06	−03	29	12	27	30	12	−05	22	−12	11	09	23
23	−17	05	03	33	42	23	12	27	24	−07	21	01	18	−06	01	12	19	−09
24	−04	05	16	22	69	01	06	15	07	12	18	18	20	40	−05	−03	−04	05
25	14	−14	10	15	40	14	−06	17	20	05	07	39	28	20	01	−13	−17	23
26	11	07	00	20	44	−02	10	37	−07	−08	03	19	23	07	−02	15	−21	−31
27	−03	29	−11	17	−07	45	08	14	24	28	−16	40	01	00	−06	10	−07	−09
28	12	51	−09	06	27	59	−11	−01	14	07	−01	−13	11	−02	−06	00	−07	−03
29	09	29	−02	46	24	43	−02	08	20	13	33	−15	06	10	01	−05	−03	00
30	18	49	−30	−07	18	53	10	13	09	07	21	11	03	−01	−12	−10	17	−06
31	10	44	−23	27	16	47	11	−01	18	09	26	26	00	−08	14	−08	−08	−03
32	22	14	−12	19	−20	40	11	10	−23	24	−17	02	06	24	06	07	24	11
33	26	−04	−02	15	−07	77	10	15	−10	25	03	−11	01	01	01	19	−10	18
34	23	10	07	03	29	02	45	30	16	02	12	15	13	20	−15	11	08	07
35	07	−15	−05	03	29	−04	69	13	12	11	05	−07	19	04	−01	−20	−17	−07
36	−11	08	03	03	−11	−12	47	11	−05	09	04	12	02	02	09	−03	07	−02
37	12	20	09	−01	36	11	−04	59	02	08	−04	06	15	21	00	−12	17	18
38	−23	−11	15	19	22	10	20	56	15	10	40	−02	08	−01	00	22	05	−12
39	−05	00	−06	10	23	20	03	49	11	41	−17	09	−03	27	07	−24	06	−29
40	08	19	02	10	06	02	32	36	25	06	20	25	11	34	10	−09	−31	06
41	12	08	02	14	15	08	08	57	22	06	16	09	22	23	−17	04	−07	−17
42	−17	22	−09	34	07	14	11	28	13	00	35	−03	25	32	−06	−04	−15	−24
43	13	14	03	−02	17	15	−01	50	30	23	03	16	09	44	−14	23	−04	−02
44	00	−10	52	22	05	10	14	11	44	−17	06	15	−01	28	−05	−09	14	02
45	−07	00	12	07	13	05	08	50	48	−22	22	08	03	08	−13	20	14	−12
46	−15	01	07	31	−08	−09	−16	08	43	01	−07	−01	27	29	15	−22	08	−08
47	16	−03	−02	19	36	17	28	14	20	−11	20	02	14	51	04	22	11	−29
48	27	16	07	25	16	−10	33	22	39	−17	22	18	09	12	06	06	−07	05
49	07	−11	14	19	−01	22	18	04	−24	60	13	−08	−10	−24	−08	−15	15	−02
50	−35	−06	12	05	−06	−06	−05	04	19	−09	01	13	04	62	25	−08	−08	09
51	−08	−02	19	13	17	05	25	02	−14	48	19	13	−21	−25	02	20	04	−12
52	02	00	06	04	02	14	04	13	−19	50	05	−33	−09	−08	−03	14	−18	18
53	05	33	18	27	00	20	02	14	−22	−01	04	08	05	00	04	35	08	01
54	26	09	−02	−11	−07	28	12	10	−21	29	35	−08	16	−04	05	02	−06	−15
55	−18	09	08	52	16	02	07	05	−04	10	31	14	37	17	−18	−04	00	−01
56	−08	28	14	19	27	12	15	−04	23	16	42	27	04	28	17	04	−03	−05
57	−02	05	06	35	17	−02	−07	20	−04	−10	54	22	27	07	−26	−18	−06	−07
58	09	03	28	02	−03	−13	13	08	35	13	53	01	07	01	07	05	16	05
59	−19	−03	01	−10	−02	17	04	10	−09	−05	03	61	39	−04	17	−09	05	−03
60	15	11	−11	12	08	11	−08	16	23	13	10	62	25	−05	−19	−06	−10	01
61	03	03	−14	25	22	02	−01	13	−13	−17	26	59	19	03	10	−01	−09	06
62	15	−12	−16	15	09	01	01	00	16	06	36	45	34	06	15	10	14	−14
63	−01	04	18	03	10	14	01	05	34	16	04	29	69	03	−13	05	00	13
64	03	−01	20	22	03	11	30	12	07	04	16	04	38	09	22	06	−13	18
65	−05	13	12	−12	−09	−03	05	18	13	−08	01	30	66	−03	27	03	−10	08
66	06	−12	20	26	−04	13	12	05	02	−06	−12	37	53	14	−25	−07	21	13
67	02	−04	10	32	13	23	−04	01	04	−07	07	45	62	07	01	09	01	−04
68	−08	16	−15	−12	15	−02	−08	22	04	05	39	36	49	19	−11	−15	−15	−13
69	−05	12	00	08	33	06	18	−13	19	−12	31	31	35	−06	−01	−02	16	−22
70	05	−02	77	−07	05	00	08	00	12	09	12	10	−13	09	−07	−03	15	00

* Decimal points have been omitted.

bring out this fact. Thus, in the first list, for Factor G, there are three tests also loaded on Factor I and two on Factor O. The probable meaning of these systematic relationships will be discussed in connection with each factor.

In general, where such parallel significant loadings occur in two factors we may attribute them to one or another of two reasons. One possible reason is that tests intended to measure one factor include items that also indicate the other factor. This could be the sole reason, the correlation between the factors being zero. Another possible reason is that the two factors are actually correlated. Sometimes both of these reasons apply. We shall attempt to decide when either impure tests or correlated factors are involved. More often than not, we shall find that even when two or more tests have significant loadings on a pair of factors, other tests have zero, or near zero, loadings on the one factor and not on the other. This may be taken to indicate that the psychological factors are orthogonal.

In other words, *common significant loadings in two or more tests is not sufficient evidence for correlation between factors.* We must often look to what we regard as insignificant loadings to provide the indications of correlation. If none of the tests with significant loadings on a factor has a zero loading on some other factor, the two factors may be correlated. For example, Tests 1–4 have significant loadings on Factor G (see Table 2) and their loadings on Factor C_1 are −.14, −.16, −.14, and −.17, respectively. Although all these loadings on C_1 are insignificant, and even quite small, by their unanimity they indicate that axis G should be rotated further away from C_1 to the extent of about 21 degrees. The angular separation of axes G and C_1 is then 111 degrees. The cosine of 111 is −.36, which indicates a correlation of −.36 between Factors G and C_1.

We shall use this rough method to obtain some idea of the amount of correlation between pairs of factors, starting from the orthogonal solution. Since the number of apparently nonzero correlations is very small (of the 105 pairs of factors, some 15 seem to be correlated more than .10), this method will serve fairly well. For more complete and accurate estimates, an oblique rotational solution would be needed. We have previously expressed some reservations concerning estimates of factor intercorrelations derived from oblique rotations. The same reservations apply to some extent to our own estimates. The chief reason is that when complete orthogonality between two factors does not occur, this may be due to our inability to prepare the appropriate tests, each as free as possible from variance in the other factor.

Tests 1–5 were composed of items scored for G. Tests 1–4 appear in this list as expected, but Test 5 (Quickness) has a loading of only .18. Thus, the quick-acting

G. General Activity

Test No.	Test title	Loadings in the Factor	Loadings in Other Factors
1	Rapid Pace	.47	
3	Energy	.44	(R −.36)
17	Personal Strengths and Weaknesses	.44	(I .44)
2	Drive for Activity	.42	
18	Feeling of Adequacy	.38	(I .36) (O .30)
4	Liking for Action	.37	
50	Impulsiveness	.35	(C_2 .62)
19	Self Confidence vs. Inferiority	.34	(I .45) (O .30)

aspect of this factor seems to drop out in this analysis, leaving the other four qualities to define it. Of Tests 1–4, only Test 3 has a significant loading on any other factor, which indicates the relative uniqueness of those score variables. As expected from the relatively low intercorrelations of scores for the G variables, none proved to have more than moderate loading on G.

Tests 17–19, designed for Factor I (*confidence vs. inferiority feeling*), have significant loadings on G. Test 17 is correlated with G probably because three of its seven items make reference to good health and physical strength. While Tests 18 and 19 have no items that give any direct clue to their loadings on G, some of their items may relate indirectly to health and strength, e.g., playing best against a superior opponent. In general, for some examinees, confidence for meeting life's competitions may well imply general drive and physical fitness.

From this line of thinking we find a logical basis for a genuine correlation between G and I. The relationship indicated by the data, however, seems to be a one-sided one, with significant loadings on G for the I tests but no significant loadings on I for the G tests. We must also consider the fact that in addition to the three G tests with near-zero loadings on I, there is one I test with a zero loading on G. From these last facts we conclude that the factors are probably orthogonal.

A possible relation of G to Factor O (*objectivity*) is suggested by weak loadings for both in Tests 18 and 19, but tests of G uniformly have close to zero loadings on Factor O, and tests of O have almost zero loadings on G. Thus, O and G should be regarded as independent.

When explaining the method used to estimate a correlation between two factors above, we used Factors G and C_1 as an example. It was concluded that the correlation between G and C_1 is about $-.36$. Factor C_1 is interpreted in this analysis as being *emotionality*. Another possible name for it is "emotional immaturity." Since the loadings of tests on this factor in Table 2 relate to the opposite pole, the negative sign for the correlation of .36 means that the greater the activity level or drive level of the individual the greater his emotionality or emotional immaturity. The negative correlation is with emotional maturity. This correlation fits nicely with the activation hypothesis of emotion in general. To some extent, freedom from emotion may mean low drive for physical activity.

The only other factor with which G seems to have appreciable correlation is Ag (*agreeableness*). The correlation indicated is approximately $-.40$. An active person is more likely to be low on Ag, i.e., to have resentments and hostilities, and also contempt and disregard for others. A more agreeable, compliant, friendly person is a less active person. It has been observed clinically that persons with low Ag score are much more likely to cause overt trouble if also high in G. Apparently a person's lowness on Ag is also limited by his lowness on G.

The leading test in this list (Maintaining One's Rights) suggests that Factor A is psychologically a socially approved form of aggressiveness or self-assertiveness. It is not the aggressiveness born of frustration. To this extent, as well as for other reasons, the frustration-aggression hypothesis has serious limitations. While there may be an element of frustration involved when a person has to maintain his rights, as indicated by Test 7, other tests of Factor A do not imply this element. The aggression involved in the exhibition of Factor A is of a nonpunitive kind. The other unique test in the list (8, Self Defense) also seems to emphasize the defensive aspect of this fac-

A. ASCENDANCE VS. SUBMISSION

7	Maintaining One's Rights	.74			
28	Social Leadership	.51	(S .59)		
6	Being Conspicuous	.49		(I .41)	
30	Liking for the Limelight	.49	(S .53)		(M .30)
9	Social Initiative	.48	(S .48)		
8	Self Defense	.46			(N .37)
31	Shyness, Bashfulness	−.44	(S −.47)		
10	Fear of Social Contacts	−.40		(I −.43)	
53	Reticence	−.33			

tor, but other tests, such as 6, 9, and 28, present more positive qualities.

There is much suggestion of correlation between Factors A and S (*sociability*) in the list above. Three of the seven tests designed for S have significant loadings on A, and one more (27, Liking for Friends and Acquaintances) just barely misses. The title of Test 28 (Social Leadership) immediately implies some degree of self-assertion. Items in Test 27 make no direct statement regarding taking the initiative, but getting acquainted and making friends is undoubtedly easier where some initiative is shown. Tests 30 and 31 contain items that are in some ways parallel to items in Test 6. Only one of the A tests (9, Social Initiative) has a significant loading on S, but Test 10 (Fear of Social Contacts) barely misses.

In spite of all these secondary relationships, four of the A and S tests have essentially zero loadings on the other factor. Tests 7 and 8 have zero loadings on S, and Tests 32 and 33 have zero or near-zero loadings on A. Thus, the factors are probably orthogonal. In distinguishing between Factors A and S psychologically, Tests 7 and 8 should accordingly be given most weight for A, and Tests 32 and 33 for S.

From the standpoint of methodology, we have here a striking example of how impure tests can give the erroneous appearance of correlated factors. If either Tests 7 and 8 or 32 and 33 had been absent from the analysis, the conclusion would have been that the factors *are* correlated. This kind of outcome probably occurs on many other occasions in which there is also actually much less correlation between the psychological factors than appears.

Another apparent factor correlation must be examined. Two A tests have second loadings on Factor I (*confidence*). Items in Test 6 make no direct reference to confidence, but actions suggested in two of them would seem to require it, e.g., "Do you usually speak out in meeting to oppose some one who you feel sure is wrong?" All the items in Test 10 refer to actions for which a feeling of confidence would be needed or would be desirable; actions such as soliciting funds, selling, making application in person, and "having it out" with a critic. The nature of items in Tests 6 and 10 thus seems to account for their loadings on Factor I. There is other evidence of correlation, however.

Although one A test has a near-zero loading on I, and vice versa, a plot of the significantly loaded tests on axes A and I shows that each axis could well be rotated slightly toward the other. The net result is a separation of 65 degrees, indicating a correlation of +.42. It may be that better item writing would succeed in reducing this correlation. This remains to be seen. On the other hand, a linking of self-confidence and self-assertiveness is logical.

Tests 11–16 and variable 70 were designed as measures of Factor M, and all have significant loadings on it. The three leading variables being what they are,

M. MASCULINITY VS. FEMININITY

70	Male Sex Membership	.77			
13	Masculine Vocational Interests	.72			
14	Masculine Avocational Interests	.64			
11	Fearfulness	—.60	(D .30)		
44	Emotional Excitability	−.52	(C_1 .44)		
16	Sympathy	−.48			
15	Disgustfulness	−.45			
12	Inhibition of Emotional Expression	.39	(C_2 −.32)		
30	Liking for the Limelight	−.30		(A .49)	(S .35)

there can be no doubt regarding its identity.

It might be surprising that more of the variables not designed primarily for M show no relationship to this factor, for any variable in which there is an appreciable sex difference should normally be expected to correlate with this factor. The reason for this will become apparent later. Briefly, the sex differences in other variables tended to be systematically confined to tests of certain factors—S, O, and Co, in particular. These relationships are accounted for in terms of small correlations of M with these factors.

Only two isolated non-M variables have significant loadings on the factor: 44 (Emotional Excitability) and 30 (Liking for the Limelight). Items or total scores indicating emotional expressiveness or excitability have previously correlated negatively with masculinity. How much this may be due to reluctance on the part of the male to admit emotionality and how much is genuine sex difference in emotionality we do not know. Two M tests, 11 (Fearfulness) and 16 (Sympathy) have loadings of .30 and .28, respectively, on Factor D. There are not enough systematic relationships between M tests and C and D tests to indicate intercorrelations of M with those factors.

It is difficult to make an accurate estimate of the correlation between Factors M and S (*sociability*). It might be as small as −.24 and as high as −.44. The loading of −.30 on Test 30 (Liking for the Limelight) may be inflated somewhat by overmodesty on the part of the males in responding to this kind of item or reluctance to admit as much exhibitionism as they feel. The males may also be overrating their bashfulness or shyness, which they indicate to the extent of a loading of −.23 on M for Test 31 (Shyness, Bashfulness). Discounting the loadings on these two tests, the correlation between M and S may be nearer to −.24 than −.44.

The correlation between M and O (*objectivity*) could be anywhere between zero and .52, depending upon where one prefers to draw the axes. Whatever the correlation, it is positive rather than negative. That is, the more masculine the person, the more objectivity, a conclusion that males would be quite prepared to accept. On the other side of the picture, females are more hypersensitive and inclined to take things personally. If the factors are not actually correlated, it will take much revision in item writing for both to eliminate sex differences in O items and to eliminate O variance in the M items. The chances are that the factors are positively correlated to some degree.

The correlation between M and Co (*cooperativeness*) is much more clearly estimated to be about −.12. From this fact the implication is that males are slightly more inclined to be fault-finding and complaining than females. The correlation is so small, however, as to be of little consequence.

Something should be said regarding the

unique status of Factor M. Comparatively speaking, it is not a stable, unitary variable. An M factor is defined in any particular analysis by the experimental variables in which there are substantial sex differences. Thus, it may be expected that a masculinity-femininity factor will differ from one analysis to another, depending upon which M-factor indicators are present. One implication of this fact is that the correlation of M with other factors will differ from one analysis to another.

Another implication of practical importance is that the nature of MF scores obtained from published tests will vary. Such a score is obtained from the *Minnesota Multiphasic Personality Inventory*, the *Terman-Miles Attitudes-Interests Analysis Test*, and the *Strong Interest Blank*, as well as from *Inventory of Factors GAMIN* and the *Guilford-Zimmerman Temperament Survey*. It can be predicted that MF scores from different tests will not intercorrelate as high as might be expected of reliable scores of the same factor. The *MMPI* MF score would presumably be weighted relatively more by sex differences in pathological symptoms. The M scores in the authors' inventories are weighted relatively more heavily in terms of interests and certain emotional qualities. The Terman-Miles MF score would involve the same variables plus additional ones derived from inkblot and word-association tests. In the interpretation of any MF score, therefore, it is important to keep in mind the areas of sex differences represented.

Tests 17–21 were designed to measure Factor I. All appear in the list above, along with twice as many others that were not so designed. Let us consider first the properties of this factor as indicated by its largest loadings.

The leading variable (Feeling of Acceptance), with no significant secondary variance, indicates that this factor is best defined as a feeling of acceptance and security in social relations. The appearance of variable 55 (Egocentrism), with negative loading, was unexpected, since it was designed as a measure of Factor O (*objectivity*). Two of the eight items in that test touch upon inferiority feelings in a projective manner: "Are you continually comparing yourself with other people?" and "Are you as quick as other people to pay compliments where they are deserved?" But it would hardly seem that this small proportion of items alone would swing the Egocentrism score so far toward I. As we shall see, there is a small correlation between I and O, and hence some basis for accepting an egocentric aspect to Factor I. On the basis of clinical observations, too, we should expect self-centeredness to be an

I. Confidence vs. Inferiority Feelings

21	Feeling of Acceptance	.64			
55	Egocentrism	−.52	(O −.31)		(Co −.37)
29	Social Poise	.46	(O .33)	(S .43)	
19	Self Confidence vs. Inferiority	.45	(O .30)	(G .34)	
17	Personal Strengths and Weaknesses	.44		(G .44)	
20	Discontent with Status	−.43			
10	Fear of Social Contacts	−.43			(A −.40)
6	Being Conspicuous	.41			(A .49)
57	Ideas of Reference	−.35	(O −.54)		
18	Feeling of Adequacy	.34	(O .30)	(G .38)	
42	Feelings of Guilt	−.34			(C_2 .32)
23	Restlessness	−.33		(N −.42)	
67	Suspiciousness of Hypocrisy	−.32		(Co −.62)	(Ag −.45)
46	Emotional Immaturity	−.31			(C_1 .43)
22	Tense-excited vs. Calm-relaxed	−.30	(O −.30)	(N −.40)	

aspect of inferiority feelings. The mechanisms relating confident behavior to acceptance and security on the one hand and to absence of egocentrism on the other are common clinical conceptions.

The long list of non-I tests significantly related to Factor I and the many significant secondary loadings for I tests suggest that there are many real linkages between I and other factors. In general, this is true, but there are other things to be considered.

There is many an opportunity for non-I score variables to be loaded on I because of the peculiar psychological status of this factor. Many an item in non-I tests pertains to quality or to behavior that has implications of strength or weakness on the part of the examinee. If he gives himself the benefit of the doubt on such items and does it systematically, and if, as is probable, some tests feature this kind of item more than others, correlations with I will be generated. If most or all of the tests for a certain non-I factor are of this kind, a correlation between that factor and I will be indicated. If only a few of the tests for a non-I factor are so affected, the factors can still be uncorrelated, but only these few tests will be loaded on the factor. Tests in the list above that seem to fall in this category are 42 (Feelings of Guilt), 67 (Suspiciousness of Hypocrisy), and 46 (Emotional Immaturity). The significant loadings on I for these tests have no implications for correlations of factors. The systematic relation of three G tests to I was discussed before, and the conclusion was that the Factors G and I are uncorrelated.

The correlation of I with Factor A was reported earlier to be +.42. I also seems to correlate to smaller degrees with Factors N, S, C_1, and O, the coefficients being +.31, +.16, −.31, and +.12, respectively. All these coefficients seem to be fairly accurately estimated, and, of course, are a function of this particular test battery. If the significant loadings would suggest correlations stronger than these figures suggest, the fault is with the tests themselves. The interpretation of these correlations would be that the confident person is inclined to be calm and composed, sociable, emotionally mature or emotion-free, and objective. With feelings of inferiority go nervousness, seclusiveness, emotional immaturity, and hypersensitiveness. Few would doubt the reasonableness of these relationships.

Variables 22–26 all lead in the list of tests related significantly to Factor N. Because they were designed for N, and because the term "nervousness" appears in the name of the leading variable, this factor has been well named. The tense vs. calm aspect, however, is not as strong as in previous results, the loading in Variable 22 (Tense-excited vs. Calm-relaxed) being only −.40.

Because of paired, significant, secondary loadings on I, D, O, Ag, and C_2, we should

N. Calmness, Composure vs. Nervousness

24	Nervousness and Jumpiness	−.69	(C_2 .40)		
26	Fatigueability	−.44		(D .37)	
23	Restlessness	−.42			(I −.33)
22	Tense-excited vs. Calm-relaxed	−.40		(O −.30)	(I −.30)
25	Annoyance and Irritability	−.40			(Ag −.39)
8	Self Defense	.37		(A .46)	
37	Emotional Depression	−.36		(D .59)	
47	Absentmindedness	−.36	(C_2 .50)		
68	Victim of Hard Luck	−.33	(Co −.35)	(O −.30)	(Ag −.31)

consider possible relations of N to those factors. The correlation of .31 between N and I has already been mentioned. The plots in planes involving N and each of the others in turn indicate essentially orthogonal relationships except between N and D.

The other apparent relationships can probably be accounted for in terms of impure tests. The two tests having Ag variance as well as I variance, for example, contain some telltale items. Some of the items in Test 25 (Annoyance and Irritability) probably pertain to things that arouse the ire of persons who are predisposed to be belligerent (low on Ag). They refer to "piling on" work, telling the driver how to drive, crowding ahead in line, and "people who hurry you." Test 69 (Victim of Hard Luck) divides its common-factor variance about equally among four factors. It has no obvious references to composure or to nervousness to account for the significant loading on N. Three of the much stronger Ag tests, however, have zero loadings on N and four other tests with significant loadings on N have essentially zero loadings on Ag. Thus, orthogonality for N and Ag is fairly well indicated.

The correlation between N and D is estimated to be −.44, with a possibility that it is nearer zero than that. The negative sign means that depression goes with nervousness, and cheerfulness goes with calmness and composure. We can say that the direction of the correlation is clearly indicated and that it is psychologically reasonable. The degree of correlation is still in doubt.

Seven tests, 27 through 33, were designed to measure sociability and all seven of them are in the list above. There is only one test in the list that was not designed for this factor—the A variable, Social Initiative.

The two variables without significant secondary loadings here are 33 (Liking for Social Affairs) and 32 (Gregariousness), the former leading all the other variables by a substantial margin. These relatively pure tests suggest that the factor should be defined as social interest. The other S variables support this definition indirectly, but after eliminating their A components they leave something that should be considered toward the defining of S. They suggest a somewhat broader definition than social interest and justify the more general term "sociability." The opposite pole would be seclusiveness, not just lack of social interest.

From the significant secondary loadings, the only factor with which S seems correlated is A. In earlier discussion it was pointed out that this is only an apparent correlation. It is all too obvious that too many S tests are impure, leaning heavily toward Factor A.

Although not apparent in the table above, there is a correlation of S with M, which has been noted in the discussion of the latter factor. The estimate was between −.24 and −.44. We also find a correlation of −.21 between S and R. This indicates that the social person is likely to be on the happy-go-lucky, carefree side of R and that the restrained person is likely

S. SOCIABILITY

33	Liking for Social Affairs	.77			
28	Social Leadership	.59	(A .51)		
30	Liking for the Limelight	.53	(A .49)		(M .30)
9	Social Initiative	.48	(A .48)		
31	Shyness, Bashfulness	−.47	(A −.44)		
27	Liking for Friends and Acquaintances	.45		(Ag .40)	
29	Social Poise	.43		(I .46)	(O .33)
32	Gregariousness	.40			

to be less sociable. Thus are the introvert and extravert poles of these two dimensions found to be in the alignment that should be expected.

With seven a priori tests of S in the analysis, we should not have been greatly surprised to find them separating into two or more groups. Layman (10), for example, found four factors in the social area: "social initiative," "gregariousness," "social inadequacy," and "social aggressiveness." The last may be aligned with our Factor A. In some respects her factor "Social inadequacy" resembles our Factor I, but Layman also reported another factor "feeling of inferiority," which comes closer. Mosier reported another social factor "platform shyness" (12), which has no counterpart in our analysis, but which may be related to the negative pole of Factor A.

In our analysis there seemed to be sufficient opportunity for "social initiative" to separate from S, as well as from A. It is doubtful whether "platform shyness" could have done so, although this, too, was a possibility. Both Layman and Mosier analyzed single items whereas we analyzed scores from groups of items. It is possible that in our groupings we have submerged one or two social factors. Only further analyses will clear up this matter.

T. REFLECTIVENESS

35	Liking for Serious Thinking	.69			
36	Analysis of Self and Others	.47			
34	Meditativeness	.45	(D .30)		
48	Daydreaming	.33		(C_1 −.39)	
40	Loneliness	.32	(D .36)		(C_2 .34)

The three variables designed to measure Factor T (34–36) lead this list. So far as our present information goes, we have little basis for suggesting any important change in this factor. It emphasizes a liking for reflective and meditative thinking, which does not preclude a liking for or indulgence in daydreaming.

We are unable to say whether this factor is distinct from a factor called "interest in thinking" found in a comprehensive analysis of interests (4) and also distinct from another factor called "autistic thinking" in the same analysis. Factor T does not seem to be an interest in thinking that is as rigorous or logical as in the former instance or as wishful as in the latter instance. Until there is evidence to the contrary, it is safe to conclude that T is not the same as either of those factors. It is clearly in line with what French (1) calls "interest in philosophizing." It may be suggested that these factors represent attitudes toward three somewhat different kinds of thinking and that there may be still other such thinking-attitude factors that have much to do with intellectual functioning.

In the list of significant loadings above, there is no indication of correlation between T and any other factor. Examination of Table 2, however, and a plotting of pertinent loadings, show that a correlation of about +.21 is called for between T and D. The estimate seems to be rather clearly determined. There is thus a probable connection between interest in reflective thinking and depression and between cheerfulness and lack of interest in reflection or meditation.

Of the six D variables, 37–42, five are significantly related to Factor D. The sixth test, 42 (Feelings of Guilt) has a loading of .28, but it has three slightly higher loadings on other factors (I, O, and C_2). From clinical observations we might well expect a more substantial relation between guilt feelings and depression. This result suggests that the connection is more pathological than normal. The presence of the qualities of depression, worry, and lack of cheerfulness high in the list, however, indicates that the factor is well named.

French (1) prefers to name this factor "loneliness" because in Mosier's analysis (12) items on loneliness were more or-

D. Depression

37	Emotional Depression	.59	(N −.36)		
41	Physical Depletion	.57			
38	Worry, Anxiety	.56		(O −.40)	
43	Emotional Fluctuations	.50		(C_1 .30)	
45	Emotional Perseveration	.50		(C_1 .48)	
39	Cheerfulness	−.49			(R −.41)
26	Fatigueability	.37	(N −.44)		
40	Loneliness	.36		(C_1 .34)	(T .32)
11	Fearfulness	.30			(M −.60)
34	Meditativeness	.30			(T .45)

thogonal to the "cycloid-disposition" factor than were the more obviously depression items. In our analysis, however, the test Loneliness divided its common-factor variance among three factors—D, C_1, and T.

The variables intended for Factor C in this analysis divided into two groups so that a clear-cut cycloid-disposition factor did not materialize. Instead, we have Factors C_1 and C_2, to be discussed later. It can be noted here that most of the D tests have low or zero projections on both C_1 and C_2, indicating that the more strongly loaded depression tests are almost orthogonal to those two factors.

Two of the C tests, however, had their highest loadings on Factor D: 43 (Emotional Fluctuations) and 45 (Emotional Perseverations). The former would ordinarily be expected to define best the cycloid-disposition factor. The presence of these two variables in the D list illustrates one important reason for the high intercorrelation of factor scores for D and C in inventory STDCR.

Actual correlations between D and other factors have been mentioned earlier. There are two: a correlation of −.44 with Factor N and a correlation of +.21 with Factor T.

Of the six C variables, five have significant loadings on Factor C_1. But we have the curious result, mentioned above, that Test 43 (Emotional Fluctuations) no longer heads the list. In fact, it is at the bottom of the list and just over the line of significance. Its largest loading (.50) is on D and its next largest (.44) is on C_2. Although Factor C_1 seems close to the former Factor C, it lacks Factor C's distinctive cycloid character. It is true that two of the five items in Test 43 do not refer directly to emotions. They read: "Are there times when your mind seems to work very slowly and other times when it works very rapidly?" and "Are you sometimes bubbling with energy and sometimes very sluggish?" The other three items, however, very pointedly refer to fluctuations of mood.

From the nature of Test 46, particularly, but also to some extent Tests 44 and 45, one might be prompted to call this factor "emotional immaturity." French (1) lists such a factor, separate from the cycloid factor. Items in Test 46 rather obviously suggest emotional immaturity; for example: "Do you work much better when you are praised?" "When failing to have your own way, do you often resort to re-

C_1. Emotionality

45	Emotional Perseveration	.48	(D .50)		
44	Emotional Excitability	.44		(M −.52)	
46	Emotional Immaturity	.43			(I −.32)
48	Daydreaming	.39		(T .33)	
58	Unwarranted Sympathy	.35	(O −.53)		
63	Faultfinding with Human Nature	.34		(Co −.69)	
43	Emotional Fluctuations	.30	(D .50)		(C_2 .44)

sentful thinking?" and "Does it upset you much to lose in a competitive game?" The items seem to contain a "spoiled-child" core. This is not so true of items in Tests 44 and 45. Examples from Test 44 are: "Do you 'get rattled' easily at critical moments?" and "Is your own mood very easily influenced by people around you, that is, by happy people or sad people?" Examples from Test 45 are: "Do you spend much time in thinking over good times you have had in the past?" and "Have you ever been bothered by having a useless thought come into your mind repeatedly?" Variables other than 44–46 contribute still other qualitative aspects to the picture. Hence, until more is known about this factor, we propose to call it by the very general term "emotionality."

In the general area of Factors D and C it looks as if we need to resort once more to the analysis of individual items. Anyone who attempts to group items homogeneously in this area is likely to encounter many difficulties. If the grouping is inadequate, the resulting factor structure is likely to be clouded, if not distorted. While Factor D in this analysis seems to be reasonably well determined, the solution with respect to C is somewhat out of line with previous findings. Perhaps, indeed, the cycloid factor as previously found should break up along the lines found here. Our conclusions in the region of Factor C must be somewhat indeterminate.

Factor C_1 seems to correlate with four other factors. Since there are so many, none can be so accurately estimated by our single-plane method. It was reported earlier that C_1 correlates +.36 with Factor G and −.31 with Factor I. The estimate of the correlation with R is as high as −.70. This means that the person with self-restraint is very likely to be also emotionally mature. The happy-go-lucky, carefree person is likely to be emotionally immature. There can be little argument about the reasonableness of the direction of this correlation. It was a surprise to find the estimate so high, since correlations between scores for C and R in Inventory STDCR have been very low. This may mean that, after all, the location of Factor C_1 is quite different from that of the former Factor C.

Factor C_1 shows some positive correlation with C_2, which has not been named, but which seems to be a kind of emotional instability. The estimate of the correlation is uncertain. It is tentatively reported as +.28 but may be as high as +.40. We should have expected more in common between these two factors than between C_1 and R.

Only three of the six tests included to measure R held together to define this factor—Tests 49, 51, and 52. The greatest change from previous findings was that Variable 50 (Impulsiveness) vanished completely from the list, with its variance transferring mainly to Factor C_2. That it would not hang together with the other R tests was apparent from the intercorrelations. This separation might have been anticipated, for the French summary (1) gives an impulsiveness factor separate and distinct from R (which he lists as rhathymia). Whether our factor C_2 qualifies as impulsiveness, however, is another question, which will be discussed later.

R. RESTRAINT VS. RHATHYMIA

49	Carefreeness vs. Restraint	−.60			(Ag −.33)
52	Liking for Action and Excitement	−.50			
51	Seriousness vs. Unconcern	.48			
39	Cheerfulness	−.41	(D −.49)		
3	Energy	−.36		(G .44)	

Other variables that failed to represent R as expected include 53 (Reticence), which we found related to Factor A, and 54 (Rapport with the Environment), which just barely misses a significant loading on R and which has a small loading (.35) on Factor O. The relation of Test 53 (Reticence) to Factor A can probably be accounted for by its two items, which read: "Are you inclined to keep your opinions to yourself during group discussions (not class discussions)?" and "Do you dislike to talk about yourself, even to close friends?" As a matter of fact, neither item necessarily pertains to talkativeness, which *is* sometimes found loaded on R.

We should have expected the test Rapport with the Environment to be more strongly related to Factor R. The two items were: "Is it difficult to 'lose yourself' even at a lively party?" and "Do you usually become so absorbed in watching an athletic contest that you completely forget yourself?" It might be maintained that of Factors S, T, and R, the latter comes closest to Jung's original idea of introversion-extraversion. It might also be said that "rapport with the environment" is the core idea of his conception. If the latter statement is true, however, it is very doubtful that Factor R of the present analysis should be identified with Jung's concept of introversion-extraversion.

The relation of Variable 52 (Liking for Action and Excitement) to Factor Ag is probably due to one of its three items, which reads: "Do you like to play pranks on others?" It is probable that the prank-playing of the happy-go-lucky person is of the puckish variety, while that of the belligerant (low Ag) person is of the vindictive variety.

The relationship of the quality of cheerfulness to both R and D has been found in a previous study (6). Cheerfulness has a different flavor or significance when it is opposite to depression than when it is a feature of a happy-go-lucky disposition. Not all cheerfulness, obviously, carries with it an unconcern or lack of seriousness.

The loading of −.38 on Factor R for the test of Energy is interesting. It suggests that Factor R is not entirely a matter of putting on the brakes. Perhaps there is also involved a lack of motivation to act in directions that ordinarily call for brakes. Then, of course, there is the suggestion of energy expenditure in the happy-go-lucky person who likes excitement.

O. OBJECTIVITY

57	Ideas of Reference	−.54	(I −.35)		
58	Unwarranted Sympathy	−.53			(C_1 −.35)
56	Hypersensitivity	−.42			
38	Worry, Anxiety	−.40			(D .56)
68	Suspicion of Actions	−.39	(Ag −.36)	(Co −.49)	
62	Overt Aggression	−.36	(Ag −.45)	(Co −.34)	
54	Rapport with the Environment	.35			
42	Feelings of Guilt	−.35	(I −.34)		(C_2 .32)
29	Social Poise	.35	(I .46)		(S .46)
55	Egocentrism	−.31	(I −.52)		
69	Victim of Hard Luck	−.31	(Ag −.31)	(Co −.35)	(N −.33)
18	Feeling of Adequacy	.30	(I .34)		(G .38)
19	Self Confidence vs. Inferiority	.30	(I .45)		(G .34)
22	Tense-excited vs. Calm-relaxed	−.30	(I −.30)		(N −.40)

The three O variables leading this list identify the factor. All four of the O tests introduced in the analysis have significant loadings. The picture of the factor is much the same as before, except that the egocentric aspect is not so strong, Test 55 having lost the greater part of its common-factor variance to Factor I.

In general, O, like Factor I, has an unusually long list of tests related to it. It shares variances in many of the same tests with I. The explanation is probably much the same as in the case of Factor I. It is probable that many items refer to things about which examinees are "touchy" and there is a tendency to respond to those items with systematic bias. Some correlation is thus generated between such item groups (tests) and Factor O. People are often "touchy" about things concerning which they feel inferior, hence I and O would tend to be correlated with the same tests. Factors I and O, according to this hypothesis, represent two foci of generalized feeling about one's personal characteristics; a feeling that is systematically indicated in terms of relations of many tests to these two factors and of I and O tests with one another.

It was noted earlier that Factors I and O seem to be correlated to the extent of +.12. This is a surprisingly small figure, in view of the many tests in which the two factors are related. It does not account for all the correlations between I and O tests, which suggests the possibility of reducing many score intercorrelations that are due to these general biases, with improved item writing and item selection.

Another small correlation of about +.12 may exist between O and Co, indicating that the more objective person is also slightly more free from critical faultfinding and that the subjective person, who dislikes criticism himself, is more likely to criticize others. From clinical evidence one would expect this correlation to be greater.

All four of the Ag tests used in this battery are at the head of the list. The two Ag tests without significant secondary loadings indicate that the person low on Factor Ag resents domination by others and carries feelings of hostility. Also high in the list is an attitude of contempt for others and a readiness to indulge in overt aggression.

The evidence is that Ag has correlations with two other factors, G and Co. There is nothing in the list above to suggest relationship to G, but the estimated correlation is −.44. This fact was discussed in connection with the treatment of Factor G.

It is very difficult to draw a conclusion as to the correlation between Ag and Co (*cooperativeness*, *tolerance*). If we confine our attention to Test 64, which has a zero loading on Ag, and to Test 27, which has a zero loading on Co, we should conclude that the factors are orthogonal. But neither test has a very large loading on its respective factor, the projections being .40 and −.38, respectively. If we were to ignore either of them, but not the other, the consequent rotation of the axis near it would result in an angle of separation of about 70 degrees and a correlation of +.34 between axes. If we were to ignore both these tests,

Ag. AGREEABLENESS

60	Resistance to Control	−.62			
59	Contempt for Others	−.61	(Co −.39)		
61	Hostility	−.59			
62	Overt Aggression	−.45	(Co −.34)	(O −.36)	
67	Suspicion of Hypocrisy	−.45	(Co −.62)		(I −.32)
27	Liking for Friends and Acquaintances	.40			(S .45)
25	Annoyance and Irritability	−.39			(N −.40)
66	Doubting Honesty of Others	−.37	(Co −.53)		
68	Suspicions of Actions	−.36	(Co −.49)		
52	Liking for Action and Excitement	−.33			(R −.50)
69	Victim of Hard Luck	−.31	(Co −.35)	(O −.31)	(N −.33)
65	Faultfinding with Industry	−.30	(Co −.66)		

Bureau Ednl. Psy. Research
Dated

the axes would be separated at an angle of about 45 degrees, indicating a correlation of +.71. Now the correlation of the total scores for Ag and Co in the Lovell data was .63. It is probable that the correlation of factors is lower than that figure, hence the estimate of .71 seems definitely too large. From the opening remarks of this discussion, we could even conclude that the correlation between Ag and Co is as low as zero.

The last alternative seems more plausible if we examine the nature of Tests 64 (Faultfinding with Society) and 27 (Liking for Friends and Acquaintances). Why is Test 64 orthogonal to Ag while the other faultfinding tests are not? It might be that the items on faultfinding with society, unlike those on faultfinding with people and with industrial situations, refer to more abstract evils or threats against which personal attack is not so obviously implied, hence their freedom from relationship with Ag. On the other hand, it may be that Ag items generally arouse the critical reaction of the unbelligerent faultfinder as well as the more aggressive reaction of the belligerant retaliator. Items in Test 27 emphasize friendliness but not opposite to the negative pole of Co. It is not unthinkable that items of still other classes might be written so as to discriminate more sharply and strongly between Ag and Co than Tests 64 and 27 seem to do.

Of the seven tests designed for Factor Co, five lead the list above, and all seven are present. Only one test (64) escapes having secondary loadings, but another (65) comes very close to the margin. All three tests have to do with faultfinding and they reveal a discontented person who seems to find something wrong about things in general. This, of course, defines the negative end of the dimension. Suspicions of the motives and actions of others complete the unfavorable picture.

The fact that items designed for Co commonly carry implications of Ag and vice versa was noted in discussing Ag. It was concluded that the degree of correlation between Ag and Co is very much in doubt. Some other correlations with Co have been mentioned, including a minor correlation of −.12 with M and another minor correlation of +.12 with O. These results indicate that the tolerant person is more likely to be feminine and objective. Both correlations are so small as to be of little consequence.

We shall have to consider several hypotheses concerning the nature of this factor without coming to any final conclusion as to its definition or its name.

The leading quality of impulsiveness here strongly suggests that this may be the impulsiveness factor. French lists such a factor (1) and the senior author found a factor of this type in one analysis (6). But a more general aura of instability or lack of control is generated by several variables in the list. The items on absentmindedness stress inability to concentrate (two items out of three): "Does your mind wander badly while you are trying to concentrate?"

Co. Cooperativeness (Tolerance)

63	Faultfinding with Human Nature	−.68			(C_1 −.34)
65	Faultfinding with Industry	−.66	(Ag −.30)		
67	Suspicion of Hypocrisy	−.62	(Ag −.45)		(I −.32)
66	Doubting Honesty of Others	−.53	(Ag −.37)		
68	Suspicion of Actions	−.49	(Ag −.36)	(O −.39)	
59	Contempt for Others	−.39	(Ag −.61)		
64	Faultfinding with Society	−.38			
55	Egocentrism	−.37		(O −.31)	(I −.52)
69	Victim of Hard Luck	−.35	(Ag −.31)	(O −.31)	(N −.33)
62	Hostility	−.34	(Ag −.45)	(O −.36)	

C_2. (UNIDENTIFIED)

50	Impulsiveness	.62		(G .35)	
47	Absentmindedness	.51	(N −.36)		
43	Emotional Fluctuations	.44	(D .50)		(C_1 −.30)
24	Nervousness and Jumpiness	.40	(N −.69)		
40	Loneliness	.34	(D .36)	(T .32)	
12	Inhibition of Emotional Expression	−.32			(M .39)
42	Feelings of Guilt	.32		(I −.34)	(O −.35)

and "Does your mind wander badly so you lose track of what you are doing?" The implications from the variables Emotional Fluctuations, and Nervousness and Jumpiness are more obvious in this connection.

It seems apparent that something more than impulsiveness is involved. The very general term "emotional instability" might be applied, but perhaps a more univocal term could be found.

Another possibility is that this is, after all, the cycloid factor that we hypothesized at the start. While the key variable Emotional Fluctuations does have a higher loading on this factor than on Factor C_1, at the same time, it has a higher loading on Factor D than on either C_1 or C_2. The greatest weakness in this hypothesis is the fact that only one other C variable has a significant loading here—Absentmindedness—whereas four others have significant loadings on C_1.

We accept as final none of these hypotheses, but suggest, as before, that there is need for a new analysis in the general area of these emotional factors based upon the intercorrelations of items. We are now in a much better position than before to make a wise choice of items for such a study. There would be no objection, of course, to the use of very short tests, if we could be sure that each test is highly homogeneous.

DISCUSSION

We have presented each factor in turn, dealing with its properties and with its external relationships, and mentioning some of the implications. There remain some more general considerations, affecting some or all factors, that should be of interest. First, there is the question of how well the results fit or support the hypotheses with which the investigation was started. Second, there is the over-all question of the extent to which the factors are intercorrelated, and of whether any general principles emerge from those intercorrelations. Third, we shall summarize in one place the apparent properties of each factor in turn.

General fit of the data to the hypotheses. We started this investigation with the hypothesis that the 13 factors allegedly measured by the three inventories would be found by a new analysis in which every factor is represented by at least three scored variables. Although we have already pointed out places where predicted factors and their properties are verified and where they are not, let us give some attention now to the general picture of factor verification.

In terms of the number of factors verified, we may say that 12, and possibly 13, of the 13 factors that were hypothesized were actually found in the analysis. The notable exception would be that two factors emerged in the area where the cycloid-disposition factor was expected. Although neither has all the properties of that factor, one of them (C_1) has most of them.

The identification of our obtained factors with previously known factors has been largely on an intuitive basis. There are no rigorous statistical tests to rely upon when making decisions as to identities of factors. However, we can be a little more objective about the matter, using certain indications from the data themselves.

Let us consider first to what extent the test variables that were expected to have significant loadings on certain factors fulfilled those expectations. If we call Factor C_1 the verification of Factor C, we can report that only six of the 70 variables failed to show significant loadings on their respective factors. There was one such discrepancy each in the case of Factors G, D, and C, and three discrepancies in the case of Factor R. From the point of view of expected significant loadings, then, about 90 per cent of the predictions were fulfilled.

Another index of the goodness of our predictions is based upon the number of times a test variable designed for a factor had a higher loading on some other. This criterion is not clear-cut, since it is expected that many items will have relation to more than one factor. It is also expected, however, that the secondary loadings would be smaller.

For seven factors—G, A, M, N, T, Ag, and Co—there were no assigned tests having higher loadings on other factors. Factors I, S, D, and O each had one test with a higher loading on another factor. As we should expect from previous discussions, the greatest discrepancies of this kind occurred for Factors C and R. Of the six tests for Factor C, four had higher loadings on factors other than C_1. Two of these higher loadings went to Factor D, one to C_2 and one to M. Of the six tests for Factor R, three had higher loadings elsewhere. One went to Factor A, one to O, and one to C_2. Thus, altogether, only 11 of the 70 tests had higher loadings elsewhere than predicted, and the differences were usually small.

Before reaching this point, a few readers have probably thought of the cliché, "You get out of a factor analysis just what you put into it." If the term "just" is taken to mean "exactly," we have abundant evidence from this analysis, and from others, that the statement is definitely not true. We sometimes wish it were possible to come nearer to fulfilling the statement in this sense than we or others seem able to do.

If the statement is intended in a general sense, we are willing to accept it. For then it applies to an experiment as well; we get out of an experiment, also, what we put into it. In a factor analysis of memory tests we expect to find memory factors. Of course, some nonmemory factors will come out of it, too, if they "were put into it." But sophistry aside, a factor solution such as is reported here is usually a rather compelling one. It would be virtually impossible to force the rotational solution so as to bring out 14 quite different, coherent, psychological variables, with a geometric structure that looks as good as the one obtained.

The most profitable way to begin a factor-analysis investigation, where one can do so, is to set up definite hypotheses concerning what factors are present in the area of study and what their properties probably are. When there has been much prior information, as in the present study, we may make enlightened predictions of the outcome. If the predictions are fulfilled, we have supported the hypotheses. The basic logic behind our experimental design is much like that of a cross-validation study.

In the attitude to be taken toward the results, there is nothing actually *compelling* us to accept the hypotheses as true. We merely accept them with more or less confidence. This is also true of the outcome from any method of investigation. The degree of acceptance of particular hypotheses lies on a scale of confidence or of permissiveness. The scientist's own judgment as to where a hypothesis belongs on this scale depends upon his attitude toward the method and the data.

Intercorrelations of the factors. We shall now discuss briefly and generally the question of intercorrelations of the factors. We are referring, of course, to the psychological factors.

In general, we rejected the use of an oblique rotational solution and the opportunity that it would have provided for estimating intercorrelations of factors. The degree of obliqueness between two primary axes is very dependent upon what particular tests of the factors one happens to have in the battery. There are many instances in which tests for one factor show systematically some affiliation with another factor because one or more items are slanted in that direction. Many times, in spite of this situation, which in an oblique solution would probably be taken to indicate correlated factors, one or more tests of the one factor still have orthogonal relations to the other. Such factors may be regarded as uncorrelated.

It was interesting to find that certain factors more than others tended to have significant loadings in two or more tests of certain other factors, or vice versa. One of these was Factor I (*confidence vs. inferiority feelings*), which had such relationships with Factors G (*general activity*), A (*ascendance*), N (*calmness, composure vs. nervousness*), and O (*objectivity*). These relationships could be attributed in part to small correlations between Factor I and others, and in part to items slanted toward I, or items in I tests slanted toward other factors. It is probable that many items were slanted toward I because they pertain to strengths and weaknesses of the individual. The examinee's appraisal of himself in such items one at a time tends to be consistent with his appraisal of himself on the items more frankly and directly dealing with Factor I; hence the correlations.

A similar hypothesis probably applies to Factor O, which showed an unusually large number of relationships to test variables designed for other factors, including G, I, Ag (*agreeableness*), and Co (*cooperativeness, tolerance*). The negative pole of Factor O is hypersensitivity. Many items, evidently, especially among those scored for this group of factors, probably pertain to things about which considerable sensitivity is felt. The "touchy" reaction to them tends to generate correlations with O items in which this generalized trait is featured.

Concomitance of significant orthogonal-factor loadings for two factors in several tests is not necessarily an indication that the factors are correlated. The indicator of factor correlations to which we paid most attention involves first steps toward oblique rotations that start where our orthogonal rotations left off. Using the rotated factor loadings of Table 2, we plotted the pertinent test vectors in all planes where nonorthogonality was suspected. Where oblique rotations seemed to be called for, we determined the angle of separation that this would entail, and from this an estimate of the correlation of the oblique axes. The method is admittedly rough, but it may serve to show where nonzero correlations probably exist and whether they are in the positive or the negative direction.

A relatively small number of pairs of factors proved not to be orthogonal: 16 out of a possible 105. The coefficients estimated are generally small or moderate. No general pattern or principle seems apparent; the correlations are fairly well scattered, though some factors are involved in more than their share of correlations. The correlations indicated for each factor may be summarized as follows:

G: small positive with C_1; moderate negative with Ag.[11]
A: moderate positive with I.
M: small positive with O; small negative with S and Co.
I: moderate positive with A; small positive with N, S, and O; small negative with C_1.

---

[11] A low correlation is defined as below .40 and a high one as .70 and above. The positive pole of each factor is as named in the list to follow.

N: small positive with I; moderate negative with D.
S: small positive with I; small negative with M and R.
T: small positive with D.
D: small positive with T; moderate negative with N.
C_1 small positive with G and C_2; small negative with I; high negative with R.
R: small negative with S; high negative with C_1.
O: small positive with I and Co.
Ag: moderate (?) positive with Co; moderate negative with G.
Co: small positive with O; moderate (?) positive with Ag; small negative with M.
C_2: small positive with C_1.

New descriptions of factors. Those who have dealt with measurements of the 13 factors of the Guilford inventories previously will be interested in the revised conceptions of the factors as indicated by the analysis. The factor descriptions are summarized here. In each case, unless otherwise specified, the high-scoring individual is described.

G. *General activity:* Energetic, rapid-moving, rapid-working person, who likes action and may sometimes be impulsive.

A. *Ascendance:* The person who upholds his rights and defends himself in face-to-face contacts; who does not mind being conspicuous, in fact may enjoy it; who through social initiative gravitates to positions of leadership; who is not fearful of social contacts; who is not inclined to keep his thoughts to himself. There is little to indicate that "submission" accurately describes the negative pole, as was formerly believed.

M. *Masculinity vs. femininity:* Has masculine interests, vocational and avocational; not emotionally excitable or expressive; not easily aroused to fear or disgust; somewhat lacking in sympathy.

I. *Confidence vs. inferiority feelings:* Feels accepted by others, confident, and adequate; socially poised; satisfied with his lot; not self centered.

N. *Calmness, composure vs. nervousness:* Calm and relaxed rather than nervous and jumpy; not restless, easily fatigued, or irritated; can concentrate on the matter at hand.

S. *Sociability:* Likes social activity and contacts, formal or informal; likes positions of social leadership; has social poise; not shy, bashful, or seclusive.

T. *Reflectiveness:* Given to meditative and reflective thinking; dreamer; philosophically inclined; has curiosity about and questioning attitude toward behavior of self and others.

D. *Depression:* Emotionally and physically depressed rather than cheerful; given to worry and anxiety and to perseverating emotions and changeable moods.

C_1. *Emotionality:* Emotions easily aroused and perseverating, yet shallow and childish; daydreamer. (Not identical with Factor C.)

R. *Restraint vs. rhathymia:* Self restrained and self controlled; serious minded rather than happy-go-lucky; not cheerfully irresponsible.

O. *Objectivity:* Takes an objective, realistic view of things; alert to his environment and and can forget himself; not beset with suspicions, hypersensitivity, unwarranted sympathies, anxiety, or feelings of guilt.

Ag. *Agreeableness:* Low-scoring individual is easily aroused to hostility; resists control by others; has contempt for others; and may be aroused to aggressive action. High-scoring person is friendly and compliant.

Co. *Cooperativeness, tolerance:* Low-scoring person is given to critical faultfinding generally; has little confidence or trust in others; self centered and self pitying.

SUMMARY AND CONCLUSIONS

The purpose of this investigation was to re-examine the potential evidence for and against the existence of the 13 factors purported to be measured by the Guilford-Martin personality inventories. The factors had originally been discovered by analysis of intercorrelations of single items, some of the results of which had never been published. Two analyses, by Lovell and Thurstone, based upon the 13 total scores from these inventories, had been interpreted to cast doubt upon the existence of the 13 factors.

In the present study, Lovell's basic data were again analyzed, but the intercorrelated variables were part scores rather than total scores. At least three relatively homogeneous short tests were made up of items scored for each of the 13 hypothesized factors, making 69 score variables altogether. A 70th variable of sex membership was added. Eighteen centroid factors were extracted and the axes were rotated orthogonally. The conclusions are:

1. The number of dimensions needed to account for the intercorrelations is at least 14, the number that could not be considered as residual factors.

2. The close agreement of the results to the hypotheses is indicated by two criteria:

a. Only six of the 70 test variables failed to have significant loadings (.30 or higher) on the factor for which they have been scored previously in the inventories.

b. Only 11 of the 70 variables had higher loadings in factors other than those for which they have been primarily scored in the inventories.

3. Thirteen factors could be interpreted as being nearly the same as those hypothesized. Some qualifications follow.

4. Interpretations of most factors were modified slightly, in most cases providing clarified pictures of the factors. Some of the greater changes were:

a. Factor R lost its impulsiveness aspect, which leaves the way open for a separate impulsiveness factor, that has been found in other analyses.

b. Factor C lost its most distinctive feature, the tendency to have fluctuations of mood. This left a factor that we have called "emotionality," but which might be more restrictively defined as "emotional immaturity."

5. The 14th factor has a strong impulsiveness aspect, but the general picture suggests a broader trait than that; perhaps an "emotional instability."

6. Intercorrelations among factors, so far as can be ascertained indirectly from the orthogonal solution, seem to be very close to zero, with relatively few exceptions.

7. Factors I (*confidence*) and C_1 (*emotionality*) show the largest numbers of non-zero correlations with other factors. Other factor intercorrelations are unsystematically distributed.

8. Partly due to factor intercorrelation and partly due to impure tests, two clusters of relationships of tests to factors were found:

a. One involves Factor I with G, A, N, O, and Co. It was hypothesized that the examinee's own evaluations of his strengths and weaknesses in these areas, particularly, have a relation to his general feeling of adequacy. A general halo or bias is thus involved.

b. Factor O is similarly involved with G, I, Ag, and Co. A similar hypothesis was proposed, to the effect that items in the latter groups pertain to things about which examinees are systematically hypersensitive. Thus another bias is produced, but it is not independent of the first.

9. A revised list of the properties of each factor is given in the general discussion.

10. It is urged that better item writing can do much to discriminate among the factors and to clarify information about their intercorrelations.

REFERENCES

1. French, J. W. *The description of personality measurements in terms of rotated factors.* Princeton, N. J.: Educational Testing Service, 1953.
2. Guilford, J. P. *An Inventory of Factors STDCR.* Beverly Hills, Cal.: Sheridan Supply Co., 1940.
3. Guilford, J. P. When not to factor analyze. *Psychol. Bull.*, 1952, **49**, 26–37.
4. Guilford, J. P., Christensen, P. R., Bond, N. A., & Sutton, M. A. A factor analysis study of human interests. *Psychol. Monogr.*, 1954, **68**, No. 4 (Whole No. 375).
5. Guilford, J. P., & Guilford, R. B. Personality factors S, E, and M and their measurement. *J. Psychol.*, 1936, **2**, 107–127.
6. Guilford, J. P., & Guilford, R. B. Personality factors D, R. T, and A. *J. abnorm. soc. Psychol.*, 1939, **34**, 21–36.
7. Guilford, J. P., & Guilford, R. B. Personality factors N and GD. *J. abnorm. soc. Psychol.*, 1939, **34**, 239–248.
8. Guilford, J. P., & Martin, H. G. *Personnel inventory.* Beverly Hills, Cal.: Sheridan Supply Co., 1943.
9. Guilford, J. P., & Martin H. G. *An inventory of factors GAMIN.* Beverly Hills, Cal.: Sheridan Supply Co., 1943.
10. Layman, E. M. An item analysis of the adjustment questionnaire. *J. Psychol.*, 1938, **29**, 540–546.
11. Lovell, C. D. A study of factor structure of thirteen personality variables. *Educ. psychol. Measmt*, 1945, **5**, 335–350.
12. Mosier, C. I. A factor analysis of certain neurotic tendencies. *Psychometrika*, 1937, **2**, 263–287.
13. Thurstone, L. L. The dimensions of temperament. *Psychometrika*, 1951, **16**, 11–20.
14. Zimmerman, W. S. A simple graphical method for orthogonal rotation of axes. *Psychometrika*, 1946, **11**, 51–56.

(*Accepted for publication October 24,* 1955)

Vol. 70, No. 11 Whole No. 418, 1956

Psychological Monographs: General and Applied

Relative Visual Direction as a Factor in Relative Distance Perceptions[1]

WALTER C. GOGEL
Army Medical Research Laboratory[2]

INTRODUCTION

It is to be hoped that perceptions of complex visual situations can be reduced to perceptual processes which are simple in the sense that their demonstration requires only simple visual situations. The present study considers two hypotheses involving visual direction and applies these hypotheses to the prediction of the apparent path of movement of an object attached to an Ames rotating trapezoidal window. The adequacy of the two hypotheses in describing the apparent path of movement of the attached object gives some evidence of the applicability of the hypotheses to more complex visual situations than the ones from which they were developed. If applicable, the hypotheses should be useful in developing new demonstrations and in understanding the perception of movement in the visual field.

The first hypothesis to be considered is concerned with binocular vision and will be called the binocular hypothesis. A statement of this hypothesis and the experimental evidence relating to it have been given in a previous study (2). In more general form, the binocular hypothesis can be stated as follows: *The geometric binocular disparity (or stereopsis) between a test object and that object which has the smallest difference in visual direction from the test object will be the most effective stereopsis in determining the apparent distance position of the test object with respect to each of the other objects in the field of view.* For convenience in applying this hypothesis, an approximation will be made. It will be assumed that the geometric binocular disparity between the test object and a particular object will determine the apparent position of the test object when it is much closer in visual direction to the particular object than to any other object in the field.

For purposes of illustrating this hypothesis, consider the schematic top-view drawing shown in Fig. 1. The line AB represents the physical position of an object which is oriented with its left end, A, nearer to the observer than its right end, B. The object AB is viewed binocularly. Because of misleading monocular cues (for example, misleading perspective cues), AB appears to the observer to be oriented in the position of the line A′B′. The line AB might represent, for example, a top view of an Ames trapezoidal window, with the small end, A, of the window closer to the observer than the large end, B, such that the window appears to be oriented at A′B′. The long, dashed, arrowed lines represent the lines of sight from a point between the observer's eyes through A (and A′) and through B (and B′). The object shown at position p_1 or p_2 of Fig. 1 is an additional binocularly

---

[1] The author wishes to thank John P. Tammaro and Kay Inaba for their help in collecting and analyzing the data.

[2] The opinions or conclusions contained in the present report are those of the author. They are not to be construed as reflecting the views or endorsement of the Department of the Army.

Bureau Ednl. Research
Dated

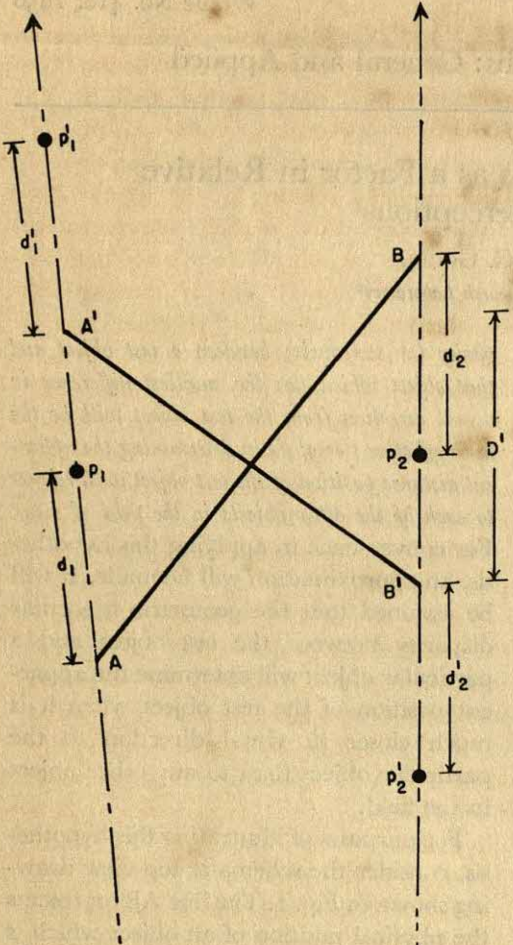


FIG. 1. Schematic representation of the difference between the physical position (p_1 or p_2) and apparent position (p_1' or p_2') of an object as a result of a difference between the physical and apparent position of another object AB.

observed object (for example, a small sphere). When this object is at p_1 it is close to the line of sight to A, and is physically a distance d_1 behind A. When this object is at p_2 it is close to the line of sight to B and is physically a distance d_2 in front of B. It will be assumed that the situation is such that only the stereopsis cue needs to be considered as a determining factor of the apparent relative distance position of the object at p_1 or p_2. In the terminology of the binocular hypothesis the object at p_1 or p_2 is a test object. A difference in visual direction as it is used in the binocular hypothesis is specified by drawing straight lines to the objects from a point between the observer's eyes. The angle formed between two of these lines measures the difference in the visual directions of the two objects. In Fig. 1, for example, when the test object is at p_1, the smallest difference in visual direction which is present occurs between this test object and A, while the largest difference in visual direction occurs between this object and B.

From the binocular hypothesis, when the test object is at p_1, the geometric binocular disparity between the test object and end A should determine the apparent depth position of the test object. The test object at p_1 should appear to be at p_1', such that d_1' (the apparent distance of the test object behind A') should be equal to d_1 (the physical distance of the test object behind A). Since d_1 is a physical and d_1' is an apparent distance, equating d_1 to d_1' requires some explanation. What is meant by $d_1'=d_1$ is that the perception of the depth difference between the test object at p_1 and the end A is not affected by the circumstance that AB appears at A'B'. In other words, the perceived distance d_1' is that which would be expected from the geometric binocular disparity between p_1 and A.

Also, from the binocular hypothesis, the apparent depth position of the test object at p_1 with respect to end B should be determined by the geometric binocular disparity between the test object and the end A. But the physical and apparent orientations of AB are not in agreement. Therefore, the physical and apparent depth between the test object at p_1 and the end B should not be in agreement. As shown in Fig. 1, the physical position of the object at p_1 is in front of the end B, but its expected apparent position p_1' is in back of the apparent position (B′) of the end B.

If the test object were placed at position p_2 instead of p_1, it would be expected from the binocular hypothesis that the object would appear to be a distance d_2' in front of B′ with $d_2'=d_2$. Also, the apparent position p_2' of the test object at p_2 would be in front of the apparent position (A′) of the left end A, while physically, p_2 is behind A. Furthermore, if two binocularly viewed test objects were present simultaneously, one at p_1 and the other at p_2, the apparent position (p_1') of the left test object would be considerably behind the apparent position (p_2') of the right test object even though p_1 and p_2 were physically at the same distance from the observer.

In the situation illustrated by Fig. 1, the binocularly observed object AB appeared at A′B′. This will be called a binocular illusion. If the physical distance is known between the test object and that part of the binocular illusion which has approximately the same visual direction as the test object, it can be predicted where the test object will be seen in relation to this portion of the illusion. Also, the apparent depth relation between the test object and other parts of the illusion can be predicted. For example, in Fig. 1 if the test object at p_1 is seen to be a distance $d_1'=d_1$ behind A′, and A′ appears to be a distance D′ behind B′, then the test object at p_1 should appear to be a distance $d_1'+D'$ behind B′.

Although the situation illustrated by Fig. 1 involves a binocular illusion, it is not meant that the binocular hypothesis applies only to illusion situations. The binocular hypothesis should also apply to the case in which the physical and apparent orientation of AB are identical. In this latter case, however, there would be little or no difference between the physical positions and the apparent positions of the test object at p_1 or p_2 with respect to either A or B and, therefore, the binocular hypothesis would not manifest itself by an error in the perceived relative position of p_1 or p_2 with respect to either A or B.

Phenomena relating to the second hypothesis are most readily demonstrated when part or all of the visual field is monocularly viewed. This hypothesis will be called the monocular hypothesis and is stated as follows: *There is a tendency for an observer to see objects as being equally distant from himself, with the strength of this tendency being inversely related to the difference between the lateral visual directions of the objects (as measured from the observing eye).* Data concerning this hypothesis can be found in a previous study (3).

Figure 1 can also be used to illustrate the monocular hypothesis. Let AB and the test object at p_1 (or p_2) be observed monocularly. Let AB appear to be oriented at A′B′ (whether or not it is physically located at A′B′ is not important). Suppose that all factors which would indicate the physical depth position of the test object are severely restricted or eliminated. The monocular hypothesis would predict that a test object located anywhere along a line of sight in the directional vicinity of A (the left, long-arrowed line) would appear to be in the depth vicinity of A′, while a test object located anywhere along a line of sight in the directional vicinity of B would be seen in the depth vicinity of B′. If a monocular test object which was physically located at p_1 were moved in a straight line to p_2, it should appear to move from A′ to B′. It will be seen that this is different from the case in which everything is viewed binocularly, since with binocular observation it would be expected that a physical movement of the test object from p_1 to p_2 would result in an apparent movement from a position (p_1') behind A′ to a position (p_2') in front of B′.

If strong factors which specify the physical position of the test object are present, the effectiveness of the tendency expressed by the monocular hypothesis will be restricted. For this reason, either or both the test object and the other objects are to be viewed monocularly, i.e., the stereopsis cue between the test object and the other objects should not be present. The adjective *monocular* in the monocular hypothesis means that strong cues to the distance position of the test object should be reduced or eliminated. But the adjective *binocular* in the binocular hypothesis means that it is essential in using this hypothesis that stereopsis cues be present between the test object and other objects in the field of view. In this sense, binocular observation is more an integral part of the binocular hypothesis than monocular observation is of the monocular hypothesis.

An Experimental Demonstration of the Two Hypotheses

In Fig. 1 the object AB extended from A to B. Therefore, if the test object were moved from p_1 to p_2, at every instant there would be some portion of the object AB which had approximately the same visual direction as the test object. This need not always occur. Consider, for example, the situation in which there is one object at A and another at B with an empty visual field in between. In this case, when the test object is moved from p_1 to p_2, the line of sight to the test object often will not be even approximately close to the line of sight to any other object. To accumulate evidence relating the two hypotheses to this type of situation and to further illustrate the hypotheses, the results from an experiment using two laterally separated playing cards are presented.

Apparatus

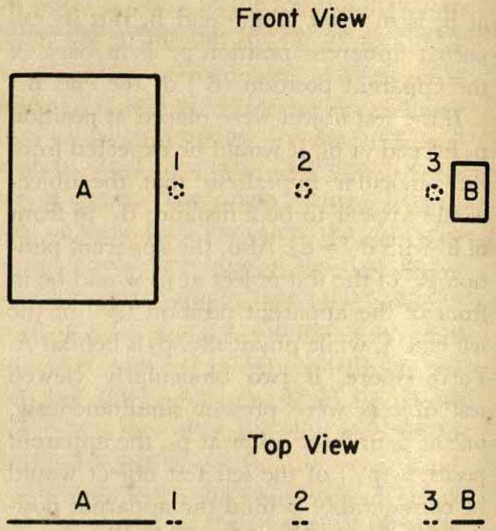


FIG. 2. Conditions for investigating the apparent depth position of a disc as a function of its lateral position between two apparently nonequidistant objects A and B.

The experimental situation used is illustrated by Fig. 2. Two seven-of-spades playing cards were presented at the same time at a distance of ten feet from the subject. The card on the left (Object A in Fig. 2) was a double-sized playing card (7 by 4 1/2 inches), and the card on the right (Object B in Fig. 2) was a half-sized playing card (1 3/4 by 1 1/8 inches). The centers of the two cards and the subjects' eyes were all at the same height from the floor. The inner edges of the two cards were laterally separated by 9 inches. A small white disc (3/8 inches in diameter) was placed between the two cards, with the center of the disc at the same height as the center of the cards. The disc was placed in one of three lateral positions. These alternative positions are indicated by the dotted lines of Fig. 2. The center of the disc was either 9/16 inch from the right edge of the left card (Position 1 of Fig. 2), 4 1/2 inches from the right edge of the left card and the left edge of the right card (Position 2 of Fig. 2), or 9/16 inch from the left edge of the right card (Position 3 of Fig. 2). The distance of the disc from the subject in all three lateral positions was the same as that of the cards. The brightness of the disc in each of the three lateral positions was equated to that of the white portions of the playing cards (2.2 foot-lamberts as measured with a Macbeth Illuminometer). No objects were visible except the disc and the two playing cards. The remainder of the room was in darkness.

Sometimes monocular, and at other times binocular, vision was used for viewing the two cards and the disc. When binocular vision was used, the disc and both of the cards were seen by the subject with both eyes. When monocular vision was used, the disc and both cards were seen by the subject with only the right eye.

Procedure

Twelve men were used as subjects. Each subject was presented with the disc in each of the three lateral positions under each of the two viewing conditions (monocular or binocular viewing). The task of the subject was to estimate in inches the apparent depth difference between the cards and the apparent depth position of the disc with respect to each of the cards. The order of presenting the various situations was varied between subjects with respect to whether the binocular or monocular view was presented first, and the order of presenting the three lateral positions of the disc.

Results

The average report of the subjects was that the right card was 18 inches behind the left card with binocular viewing, and 36 inches behind the left card with monocular viewing. The average reports involving the disc and the standard deviations of these reports are given in Table 1. In computing these summarized results, the card, with respect to which the distance of the disc was being judged, determined the zero position of the response. By this it is meant that with the disc in Position 3, for example, if one subject stated that the disc appeared 4 inches in front of the right card and another subject stated that the disc appeared 2 inches behind the right card, the average of these two reports would be that the disc appeared 1 inch in front of the right card. Each mean and standard deviation in Table 1 is determined from the reports of the same twelve subjects.

The binocular results of Table 1 involve the binocular hypothesis. Following the terminology of the binocular hypothesis, the disc was the test object. The binocular illusion was produced by the misleading size cue between the binocularly observed playing cards. The size cue was such that the small (right) card appeared farther from the subject than the large (left) card, even though both cards were physically at the same distance from the subject. From

TABLE 1

MEANS AND STANDARD DEVIATIONS IN INCHES OF THE APPARENT DEPTH POSITION OF A DISC AS A FUNCTION OF ITS LATERAL POSITION BETWEEN TWO PHYSICALLY EQUIDISTANT BUT APPARENTLY NONEQUIDISTANT PLAYING CARDS A AND B

Position of Disc	Binocular				Monocular			
	Behind A		In Front of B		Behind A		In Front of B	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
1	0	4	20	25	16	32	21	36
2	8	5	8	9	22	20	15	21
3	18	17	1	5	40	36	−4	18

Table 1, as the disc was moved from a directional position close to Card A (Position 1) to a directional position close to Card B (Position 3), the results changed from an average report of no depth difference from A and 20 inches in front of B to an average report of 18 inches behind A and 1 inch in front of B. Since the disc (the test object) was physically always at the same distance from the subject as Cards A and B, the results indicate that the physical depth position of the disc with respect to a particular card (or the geometric binocular disparity between the disc and that card) was most correctly perceived when the smallest difference in visual direction occurred between the disc and the particular card. When the disc was laterally midway between the two cards (Position 2), it appeared midway between the two cards in depth. These results suggest that a simple relationship exists between the apparent depth position of the disc and the ratio of the lateral separations of the disc from each of the cards.

The monocular results of Table 1 involve the monocular hypothesis. From the monocular hypothesis, as the difference between the lateral visual directions of the disc and a particular card decreased, the disc should have appeared to move in depth toward the apparent distance of that card. From the monocular results of Table 1, as the disc was placed laterally farther from Card A and laterally closer to Card B, the average report of the depth between the disc and a particular card increased with respect to Card A and decreased with respect to Card B. Of the three lateral positions, the position of least lateral separation of the disc from a card was the position at which the disc appeared nearest to that card in depth. One of the subjects in this monocular situation saw the disc located behind both of the cards for all three positions of the disc. If the results from this subject are removed from the monocular portion of Table 1, the mean results for this portion of the table are 6 inches, 17 inches, and 33 inches behind A for Positions 1, 2, and 3, respectively, and 28 inches, 18 inches, and 1 inch in front of B for Positions 1, 2, and 3, respectively.

With binocular vision, the apparent depth between the disc and Card A increased with each increase in the lateral separation of the disc from that card for 11 out of 12 subjects. The apparent depth between the disc and Card B increased with each increase in lateral separation of the disc from that card for 10 out of 12 subjects. The chance probability of a continuous increase in apparent depth occurring between the disc and a particular card for the two increases in lateral separation is 1/4 for a single subject. Using a p of 1/4 and a q of 3/4 in the binomial expansion, the probability that either 10 or 11 or more subjects out of 12 would see this continuous increase by chance is far less than .01. With monocular vision for 7 out of 12 subjects, the apparent depth between the disc and either card increased with each increase in the lateral separation of the disc from that card. The probability of 7 or more subjects out of 12 seeing this continuous increase by chance is less than .02.

In this experiment, the differences between the mean results from using monocular observation were similar to the differences between the mean results from using binocular observation. But, as pointed out previously, it should not be concluded that the predictions which would follow from the binocular hypothesis are always similar to the predictions which would result from the monocular hypothesis. If, for example, the disc had been placed physically more distant than the right card with lateral Position 3, it should have appeared at approximately the distance of the right card with monocular observation (the monocular hypothesis), but behind the right card with binocular observation (the binocular hypothesis).

Application to a Complex Visual Situation

The situation to which these two hypotheses were applied experimentally is the path of apparent motion of a small object attached to the Ames rotating trapezoidal window. Ames (1) found that if a window which was made in the form of a trapezoid was rotated, it appeared, with monocular observation, to oscillate back and forth. The large end of the window always appeared to be nearer than the small end, with the result that the window appeared to reverse its direction of rotation after each movement through an apparent angle of approximately 100 degrees. He also reports that a cube attached to an end of the rotating trapezoidal window would appear to leave its point of attachment and to move in a circular path. In discussing the apparent movement of the cube, Ames remarks,

> In its rotation as the cube comes toward us, its retinal image increases, and we assume it is coming nearer to us; as it goes away from us, its retinal image becomes smaller, and we assume it is going away. . . . The variations in the sizes of the retinal images and the rate of their movement across our retina, although not uniform or constant in time, are nevertheless so translated, and the cube appears to be moving in a circular path at a relatively constant speed (1, p. 18).

Kilpatrick (4) attached a playing card to the top of the small end of the rotating trapezoidal window. In one case the card was in the plane of the window and in another case the plane of the card was perpendicular to the plane of the window. Using monocular vision the card appeared in the former case to oscillate with the window, while in the latter case "the 'real' motion of the card was perceived" (4, p. 155). Kilpatrick attributes the difference in the apparent motion of the card in these two cases to observer assumptions of "togetherness" or "apartness" of the card and window.

The apparent path of movement of an object attached to a rotating trapezoidal window can be predicted from the two hypotheses with which this paper is concerned. This is considered with the aid of Fig. 3. The solid or dotted straight lines inscribed within the circles of Fig. 3 are schematic top-view representations of a trapezoidal window. A white square is rigidly attached to the large end of the window. A top-view drawing of the square would be a short straight line in, or parallel to, the long straight line representing the window. Throughout Fig. 3 the square is drawn as a square in order to distinguish it from the window. Actually, the squares in Fig. 3 represent an object whose thickness is no greater than the thickness of the window. The letters L and S stand for the large and small end of the window, respectively. The physical positions of the trapezoidal window and the square are indicated by the solid straight lines and filled-in squares, respectively. The apparent positions of the window and square for each physical position shown directly above it are indicated by the dashed lines and outline squares, respectively. The positions of the window are either those shown by Ames (1, p. 6), or are an interpolation between several of the positions shown by Ames. The left half of Fig. 3 is concerned with binocular observation of the window and square, and the right half with monocular observation of the window and square. If the position of the observer were indicated in Fig. 3, it would be shown below each top-view drawing.

A two-dimensional object can be attached to the trapezoidal window so that its visual direction with respect to each part of the window changes as the window is rotated. A situation of this type is shown in the upper half of Fig. 3. The square is

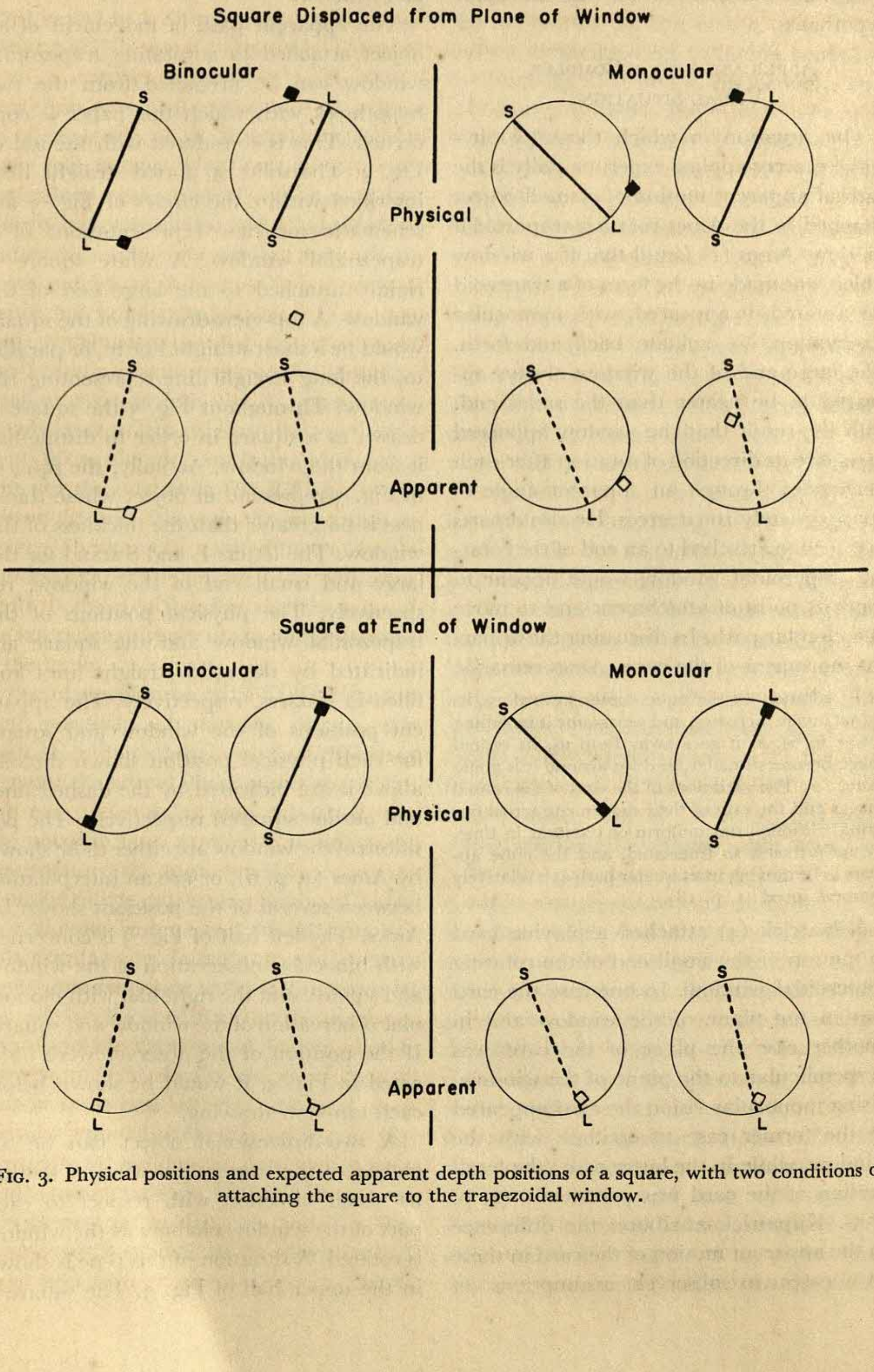


FIG. 3. Physical positions and expected apparent depth positions of a square, with two conditions of attaching the square to the trapezoidal window.

rigidly attached to the large end of the window but is displaced from the plane of the window. Or, the two-dimensional object can be attached so that, throughout the rotation of the window, there is little difference between the visual direction of the object and the visual direction of the part of the window to which the object is attached. A situation of this type is shown in the lower half of Fig. 3. The square is rigidly attached at the large end of the window and is in the plane of the window.

From the two hypotheses with which this paper is concerned, as the window rotates, the apparent path of movement of the attached square should differ, depending upon whether the part of the window which is most nearly at the same visual direction as the square is always the same part of the window or is a different part of the window at different times. Consider the upper left quadrant of Fig. 3. This quadrant involves binocular observation. In the upper left drawing of this quadrant, the large end of the window is physically in front of the small end, and the square has approximately the same visual direction as the small end. From the binocular hypothesis, in this instance there should be little or no difference between the apparent and physical position of the square with respect to the small end of the window. Since the square is physically in front of the small end by approximately one length of the window, the apparent position of the square should also be approximately one length of the window in front of the small end. But (as shown in the lower left drawing of this quadrant) the large end of the window appears closer to the observer than the small end by approximately one length of the window. Consequently, the square should appear to be in the depth vicinity of the large end. In the upper right drawing of this upper left quadrant, the square has approximately the same visual direction as the portion of the window which physically is approximately one-fourth of a window length behind the small end of the window. The square is physically a certain distance "d" behind this portion of the window. It should, by the binocular hypothesis, also appear to be distance "d" behind this portion of the window. But the small end of the window appears to be behind the large end (as shown in the lower right drawing of this quadrant). Therefore, the square, by appearing to be the distance "d" behind the portion of the window with which it has approximately the the same visual direction, should also appear to be behind the apparently far end of the window. This means that when the window rotates from the physical position shown by the upper left drawing of this upper left quadrant to the physical position shown by the upper right drawing of this quadrant, the apparent position of the square (as shown by the lower drawings of this quadrant) should move from the apparent depth vicinity of the apparently near end of the window to an apparent depth position in back of the apparently far end of the window. If the window is physically rotating counterclockwise in the top-view drawings, it will be seen that in the lower right drawing of this upper left quadrant the square is moving to the left with the small (apparently far) end of the window moving to the right. The square is about to appear to pass through the plane of the window behind the apparently far end.

The binocular hypothesis indicates that with binocular observation the attached square will be correctly perceived with respect to the portion of the window which has the same visual direction as the square. The general method of determining the predicted apparent path of movement of the square with respect to the trapezoidal window for binocular observation is as follows: (*a*) draw a top view of the apparent position of the window at a particular instant; (*b*) determine which portion of the window at this instant has most nearly the same visual direction (line of sight) as the attached object (the square); (*c*) determine the physical depth relation between this portion of the window and the square; (*d*) draw the apparent position of the square with respect to the apparent position of the window so that the physical and apparent depth relations between the square and the portion of the window most in line of sight with the square are in agreement; (*e*) repeat this procedure for successive physical positions of the trapezoidal window and attached square; (*f*) connect the successive apparent positions of the square to give the apparent path of movement of the square.

Consider the lower left quadrant of Fig. 3 which also involves binocular observation. The attached square is at the same visual direction as the large end of the window and would remain so throughout the rotation of the window. In this case, according to the binocular hypothesis, the physical depth position of the square should always be most correctly perceived with respect to the large end of the window. The large end of the rotating trapezoidal window appears to oscillate back and forth, and the square (as indicated by the lower drawings of this quadrant) should appear to move with the large end. Since the square should always appear at the same distance position as the physically large (apparently near) end of the window, it should not appear at any time to pass through the plane of the window behind this apparently near end.

With monocular observation, the physical depth position of the square should not necessarily be correctly perceived in relation to the portion of the window which has approximately the same visual direction. Instead, from the monocular hypothesis, the expected apparent depth position of the square would be near the apparent depth position of this portion of the window. With monocular, as with binocular observation, displacing the square from the plane of the window (see the upper right quadrant of Fig. 3) should make the square appear to vary its depth position in relation to the rotating trapezoidal window. But, as will be seen by comparing the monocular and binocular situations in the upper half of Fig. 3, the square in this position should not appear to move as far back in relation to the window with monocular as with binocular observation. With monocular observation and with the square displaced from the plane of the window, at some time during the movement the square should appear to pass through the plane of the window behind the apparently near end but not behind the apparently far end. This differs from the same physical situation using binocular observation since, with binocular observation, the square at some time should appear to pass through the plane of the window behind the apparently far end.

In the lower right quadrant of Fig. 3 (monocular observation), the square is physically in the plane of the window at the large end of the window. From the monocular hypothesis, when the window is rotated the square should appear to move (oscillate) with the large end of the apparently oscillating window (as indicated by the lower two drawings of this quadrant). With the square in the plane of the window, with either monocular or binocular observation, the square should not appear to pass through the plane of the window behind the apparently near end.

Apparatus

To test these predictions, a 3/4-inch white square was attached by a thin black wire to the large end of a trapezoidal window. This window was part of a kit (Portable Model, Catalog No. RTW1) purchased from the Institute for Associated Research, Hanover, N. H. The window was physically rotated in a direction such that the end of the window on the left moved toward the observer, and the end of the window on the right moved away from the observer. Four alternate positions of the square with respect to the window were used. These positions are illustrated in Fig. 4, which shows the trapezoidal window with its plane parallel to the frontal plane of the observer. In Position A in Fig. 4 the square was in the plane of the window one-half inch above the large end. This position of the square is very similar to that shown in the lower part of Fig. 3. In Position B, the square was in the plane of window three inches above the large end. In Position C, the square was in the plane of the window but was displaced three inches from the large end. In Position D, the square was displaced three inches from the plane of the window. This last position of the square is very similar to that shown in the upper part of Fig. 3. These positions of the square were presented to the subjects one at a time.

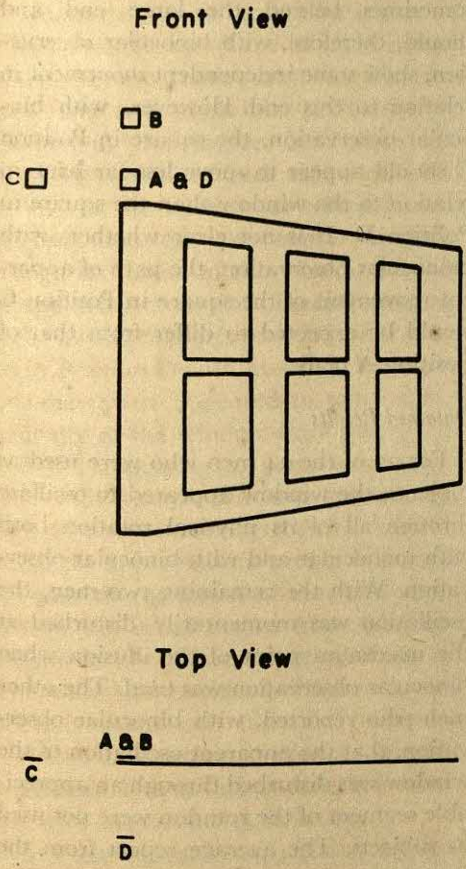


FIG. 4. Physical positions at which the square was attached to the trapezoidal window.

The center of rotation of the trapezoidal window was placed 10 feet in front of the subjects. Dark cloth was placed in back and to the right and left of the window. The window was illuminated from below and also from the two sides with low-intensity white light, so that the window and attached square could be seen throughout the rotation. Dark baffles were placed so that none of the light sources was visible to the subjects. These baffles also were not visible to the subjects. The subject looked through two small apertures located directly in front of his eyes. These apertures restricted the visibility to the vicinity of the trapezoidal window. When the subject looked through the apertures, he saw what appeared to be a window with an adjacent square, all presented against a uniform, dark gray background. The eyes of the subject were at the height of the middle of the window. When binocular vision of the window and square was required, the subject saw the window and square through each of the two apertures. When monocular vision was required, the view through the left aperture was obscured by a black opaque card, i.e., the window and square were seen with only the right eye of the subject.

The brightness of the window when the plane of the window was parallel to the frontal plane of the subject was 1.2 foot-lamberts, as measured with a Macbeth Illuminometer. From the viewing position of the subject, this brightness was reduced by placing a neutral density gelatin filter (Wratten ND .40) of approximately 40 per cent transmission in each of the viewing apertures. As in the demonstration of the rotating trapezoidal window by Ames (1), the axis of rotation was vertical and was in the plane of the window. The time required for a complete rotation of the window was 28 seconds.

Procedure

Twenty-four men were used as subjects. After preliminary observations with some of these, it was decided to have the subjects indicate the apparent path of the square by drawing a simple, top-view diagram of this path in relation to the center of the window. If the square appeared to move behind the large (apparently near) end of the window, the subjects were asked to specify the position at which the square appeared to pass through the plane of the window behind this apparently near end. Each subject was presented with the square at each of its four positions with monocular and also with binocular observation of the window and square. The order of presenting the various situations was varied between subjects with respect to whether the binocular or monocular view was presented first, and the order of presenting the four positions of the square.

Expected Results

As suggested by the discussion involving Fig. 3, for both monocular and binocular observation the square in Position A should appear to oscillate with the apparently near end of the window, and should not appear to pass through the plane of the window behind this apparently near end. In Position D, the square should appear to move independently of the window through part of its travel, appearing to pass through the plane of the window behind the apparently near end. But, with binocular observation, the square in Position D should appear to move farther back in relation to the apparently near end than with monocular observation. With Position B, the large end of the rotating window always has a smaller difference in visual direction in relation to the square than has any other portion of the window. Position B, with either monocular or binocular observation, should give results similar to those from position A. With Position C also, the large end of the window is always the portion of the window closest in visual direction to the square. But the square is physically sometimes in front and sometimes behind the large end and should, therefore, with binocular observation, show some independent movement in relation to this end. However, with binocular observation, the square in Position C should appear to move less far back in relation to the window than the square in Position D. It is not clear whether, with monocular observation, the path of apparent movement of the square in Position C would be expected to differ from that of Position A or B.

Obtained Results

For 22 of the 24 men who were used as subjects, the window appeared to oscillate through all of its physical rotation both with monocular and with binocular observation. With the remaining two men, the oscillation was momentarily disturbed at the maximum point of the illusion when binocular observation was used. The other men who reported, with binocular observation, that the apparent oscillation of the window was disturbed through an appreciable segment of the rotation were not used as subjects. The average report from the 24 subjects was that the apparent oscillation of the window extended through an angle of 100 degrees ($SD = 20°$) for mo-

TABLE 2

MEANS AND STANDARD DEVIATIONS OF THE PORTION OF THE ROTATING TRAPEZOIDAL WINDOW OVER WHICH THE SQUARE APPEARED TO PASS WITH THE FOUR CONDITIONS ILLUSTRATED BY FIGURE 4

Vision	A		B		C		D	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
Binocular	0	0	.05	.20	.22	.27	.83	.54
Monocular	0	0	.05	.20	.17	.28	.43	.24

nocular observation, and through an angle of 109 degrees ($SD=22°$) for binocular observation.

The average results and the standard deviations of these results for the apparent path of movement of the square are given in Table 2. The mean results in Table 2 are average distances expressed in units of the window at which the square appeared to pass through the plane of the window behind the apparently near end. For example, with both monocular and binocular observation, the square in Position A appeared to oscillate back and forth with the end of the window to which it was attached, and did not appear to pass through the plane of the window behind the apparently near end. This is indicated by the average reports of zero in Table 2. But in Position D with binocular observation, the square appeared to move independently of the window through part of its travel, appearing to pass through the plane of the window at an average distance of 83 per cent of the length of the window behind the apparently near end. It will be seen that the direction of the differences between the mean results from the four positions of the square are in good agreement with the predicted differences for both monocular and binocular observation. Also, as expected, the average report for Position D was larger with binocular than with monocular observation.

The *t* test was used to determine the significance of the difference between the means of Table 2. A distribution of differences was formed from the two distributions whose mean difference was to be tested for significance. The average report from binocular observation with Position D was significantly different from that with Positions A, B, or C, beyond the .001 level of confidence ($t=7.0$, 6.6, and 5.0, respectively). Also, the average report from binocular observation with Position C was significantly different from that with Positions A or B beyond the .001 and .05 level of confidence ($t=3.9$ and 2.3), respectively. The average report from monocular observation with Position D was significantly different from that with Positions A, B, or C beyond the .001 level of confidence ($t=8.6$, 7.0, and 3.9, respectively). The average report from monocular observation with Position C was significantly different from that with Positions A or B beyond the .01 and .05 levels of confidence ($t=2.9$ and 2.1), respectively. With either binocular or monocular observation, the average report from Position B was not significantly different from that from Position A at the .10 level of confidence ($t=1.2$ in both cases). The average report from Position D, with binocular observation, was significantly different from that from Position D, with monocular observation, beyond the .001 level of confidence ($t=4.3$).

Discussion

A comparison between the results from using the square in Positions A and D permits an evaluation to be made of the importance of the changes in the retinal size of the square in determining the apparent path of the square. It will be noted that practically the same series of changes in the size of the retinal image of the square occurred when the square was at either Position A or Position D. The apparent path of the square in these two positions, however, was considerably different.

A comparison between the results from using the square at Positions A and B permits an evaluation to be made of the importance of the "togetherness-apartness" factor in determining the apparent path of the square. From this factor it would be expected that the apparent path of the square in Positions A and B would differ, since the square was more separated from the window in Position B than in Position A. But this expectation is not supported by the analysis of the results.

Ames attached a three-dimensional object to the rotating trapezoidal window. Such an object attached to the rotating window so as to duplicate as nearly as possible Position A of Fig. 4 might be expected to have some differences in its apparent movement from that of the two-dimensional object used in Position A of this experiment, since a three-dimensional object extends from the plane of the window. Kilpatrick (4) attached a two-dimensional object to the rotating trapezoidal window. The playing card used by Kilpatrick behaved much like the square in Position A when the card was in the plane of the window, but not when the plane of the card was perpendicular to the plane of the window. In this latter case, portions of the card extended from the plane of the window. In both the experiments by Ames and by Kilpatrick, the object was attached to the small end of the window and, therefore, an explanation of the apparent path of the attached object when it extended from the plane of the window is probably complicated by overlay during some part of the rotation. In the present experiment, the square was attached to the large end of the trapezoidal window rather than to the small end in order to avoid having the square overlay any portion of the window during any part of the rotation.

There are some indications that factors other than those represented by the two hypotheses were operating. The average of the reports from Position D for binocular observation was not as large as expected. From the first hypothesis, the square in Position D should have appeared to move completely around (behind) the window (see the upper left quadrant of Fig. 3). Also, the average reported shape of the path of movement of the square in Position D, with monocular observation, was different from the shape which resulted when the expected path of the square was plotted using the second hypothesis. The results of this experiment, however, demonstrate that many differences between the various conditions can be successfully predicted from the two hypotheses.

Ames has indicated that, when binocular rather than monocular observation was used, the apparent oscillation of the rotating trapezoidal window was disturbed when the distance of the observer from the window was about 10 feet (1, p. 9). In the present experiment, the subjects were asked to make a simple, top-view drawing of the apparent motion of the window prior to being presented with the attached square in any of the four positions. Also, during the viewing of each of the four positions of the square, the subject was questioned about the apparent motion of the window. As mentioned previously, with 22 subjects there was no reported disturbance in the apparent oscillation of the window for either binocular or monocular observation. This difference between the results of of the two studies might be attributed to the difference in the length of the two windows used (23 1/2 inches in the study by Ames and 13 1/4 inches in the present study), or to the presence of a rectangular window above the trapezoidal window in the study by Ames but not in the present study. From the binocular hypothesis, in the study by Ames, when the ends of the rectangular window were in the directional vicinity of the corresponding ends of the trapezoidal window, there should have been some tendency to see the correct orientation of the two windows in relation to each other. Consequently, a disturbance in the apparent oscillation of the trapezoidal window might be expected.

When the physical position of the window is that shown in the upper left drawing of the upper left quadrant of Fig. 3, the amount of illusion involved in the perception of the position of the window is small. In this case, as well as for positions of the window involving larger amounts of binocular illusion, the perceived position of the square was predicted by the binocular hypothesis. If the binocular hypothesis were to be restricted so as to apply only to illusion situations, the attempt to explain the total perceived path of the attached square would involve an explanatory discontinuity as the illusion became effectively nonexistent. If only in the interest of being parsimonious, the binocular hypothesis can be considered as applying to binocular nonillusion as well as binocular illusion situations. If, for some reason, the rotating trapezoidal window had been seen in its actual (physical) orientation throughout its rotation, the attached square would have been correctly perceived with binocular observation as moving in a circular path for each of the four positions of the square which were used in this experiment. It is suggested, however, that it would move in this circular path for the same reasons that the conditions of attachment had an effect upon the apparent path of movement when the binocular illusion was present.

EQUATIONS OF PERCEIVED RELATIVE DISTANCE

From the Binocular Hypothesis

The binocular hypothesis specifies which stereopsis is most effective in determining the apparent position of a test object, but it does not indicate the degree of the effectiveness. The approximation was made that when the test object was much closer in direction to one object than to another, the stereopsis between the test object and the directionally adjacent object would determine the apparent position of the test object. As illustrated by the first experiment of the paper, however, the direction of the test object might be considerably different from the direction of each of the other objects in view. A more quantitative statement is needed than that provided by the binocular hypothesis. The following development represents a first approach to such a statement.

It was suggested from the results of the first experiment of this report that, with binocular observation (see the left half of Table 1), a simple relationship exists between the apparent relative depth position of the disc and the ratio of the lateral separation of the disc from each of the playing cards. The following notation is helpful in considering this situation:

L = physical lateral separation of the two cards,
L' = apparent lateral separation of the two cards,
x = physical lateral separation of the disc and left card,
x' = apparent lateral separation of the disc and left card,
D = physical depth difference between the two cards,
D' = apparent depth difference between the two cards,
y = physical depth difference between the disc and left card,
y' = apparent depth difference between the disc and left card.

In the experiment using the two playing cards, $D=0$ and $y=0$, it will be assumed that $L=L'$ and $x=x'$, but from the experimental results, D was not equal to D' and y was not always equal to y'. The binocular results in Table 1 suggest that the following equation applies to that portion of the experiment in which binocular observation was used: $y'=xD'/L$ (*Equation 1*). For example, when $x=L/2$, the disc appeared midway in depth between the two cards, i.e., $y'=D'/2$.

The general case will be considered in which neither D nor y are necessarily zero. This is discussed with the aid of Fig. 5. Figure 5 is a top-view drawing. If the position of the observer's eyes were indicated in this figure, it would be shown below the drawing. Points A and B represent the physical positions of two binocularly observed objects A and B. The point A′ represents the apparent relative position of A, and the point B′ represents the apparent relative position of B. Object B appears to be farther behind Object A than it actually is. This might result, for example, if A and B were similarly shaped objects with the physical size of B smaller than that of A. The point A′ and A in Fig. 5 are shown at the same position. All the error in the perception of the physically relative depth of A and B is attributed to B. This is not a necessary assumption but is made for convenience. The line connecting A and B and the line connecting A′ and B′ (or A and B′) are drawn for purposes of discussion. The point labeled "C" represents the physical position of a binocularly viewed Object C which has a different shape than A or B, so that the size cue does not occur between it and either A or B. This difference in shape is suggested by the square which is used to identify Object C. The apparent relative position of Object C is labeled C′. Two imaginary axes X and Y are placed perpendicular to each other. These axes are represented by the long solid lines. The X axis is parallel to the frontal plane of the observer and the Y axis extends away from the observer in both actual and apparent depth. The location of the intersection of these two axes is arbitrary. For convenience, it is located in Fig. 5 at Point A (or A′). Point C has the coordinate values x, y. Point C′ has the coordinate values x', y', or since x is assumed to equal x', C' has the coordinate values x, y'. The distance t represents the physical distance between C and the line AB. The distance t' represents the apparent distance between C' and the line $A'B'$. Both t and t' will be measured from C and C' in a direction perpendicular to the x axis. This simplifies the resulting equation. Actually, t or t' and the ordinate values of the various points should be measured along imaginary straight lines connecting the various points to the observer. But, since the distance of the observer to A, B, or C is usually considerably greater than the coordinate values of either A, B, or C as represented in Fig. 5, the above approximation should give reasonably accurate results.

In order to consider the general case, an additional assumption is required. This assumption is that for a particular physical and apparent depth relation between A and B, $t = t'$ for all values of y for any particular value of x. There is some evidence which bears upon this assumption. In a previous study (2), the binocular illusion was produced by a row of nine similar objects of differing size. Since more than two objects were used, the space between the end objects contained additional similar objects. The results of this previous study were generalized to mean that the depth position of the binocular test object was correctly perceived with respect to that object of the binocular illusion which had approximately the same visual direction as the test object. In the situation represented by Fig. 5, only two objects, A and B, are used to produce the binocular illusion. Except for the test object, C, the spatial interval between these two objects is empty. But from the binocular results of Table 1 it will be seen that, in this type of situation, the test object was correctly perceived with respect to that part of the line AB which had the same visual direction as the test object. From the binocular results of Table 1, when the disc was placed at the three different physical positions along a straight line connecting Cards A and B (see Fig. 2), it appeared to be placed along a straight line connecting the apparent positions of A and B. It might, therefore, be generalized with respect to Fig. 5 that the physical depth position of the binocular test object C will be perceived as though it were correctly seen in relation to that point on the imaginary straight line $A'B'$ which has approximately the same visual direction as the test object, i.e., $t = t'$.

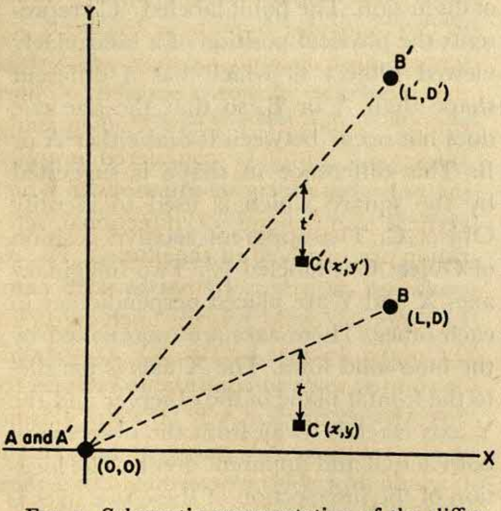


FIG. 5. Schematic representation of the difference between the physical position (C) and apparent position (C') of an object as a result of a difference between the physical and apparent relative positions of two other objects A and B

The equation giving the apparent depth position (y') of the test object (C) as a function of its physical depth position (y), its physical lateral position (x), the slope (m) of AB, and the slope (m') of $A'B'$ is $y' = y + x(m' - m)$. This is developed as follows (see Fig. 5):

Y = ordinate value of any point along AB,
Y' = ordinate value of any point along $A'B'$,
$m = D/L$ = slope of AB,
$m' = D'/L'$ = slope of $A'B'$,
t = physical distance perpendicular to the x axis between C and the line AB,
t' = apparent distance perpendicular to the x axis between C' and the line $A'B'$.

The Y value for a value of x is given by $Y=mx$. The Y' value for a value of x' is given by $Y'=m'x'$.

But,

$$Y=t+y,$$

and,

$$Y'=t'+y'.$$

Assume,

$$t=t'.$$

Therefore,

$$Y-y=Y'-y'$$

and,

$$mx-y=m'x'-y'.$$

Assume,

$$x=x'$$

Therefore,

$$mx-y=m'x-y'$$

or,

$$(\textit{Equation 2})\ y'=y+x(m'-m).$$

Equation 2 contains some terms which represent physical position and others which represent apparent position. Apparent position is expressed in physical terms. For example, a measured apparent depth of 12 inches between two objects means that the apparent depth is the same as that which would occur when observing two objects, one 12 inches behind the other, in a situation where many cues to their physical relative distance are available. Equation 2 will be illustrated by an example. In Fig. 5, suppose that the imaginary line AB forms a 30-degree angle with the frontal plane of the observer ($m=\sqrt{3/3}$) but appears as though it subtended a 60-degree angle ($m'=\sqrt{3}$). Suppose, also, that Object C is located in depth 6 inches in back of Object A ($y=6$ inches) and is laterally 10 inches to the right of A between A and B ($x=10$ inches). The apparent distance position of C is given by $y'=6$ inches $+10$ inches $(\sqrt{3}-\sqrt{3/3})$ or $y'=17.5$ inches. With binocular observation, Object C should appear as though it were physically located approximately 18 inches behind Object A and were correctly perceived in that position.

The previous equation $y=D'x/L$ (Equation 1) or $y'=m'x$ is a special case of Equation 2 which occurs when $y=0$ and $m=0$. These are the conditions for the first experiment of this report.

There are limits to the values of y, x, and m beyond which Equation 2 cannot be expected to hold. It cannot be expected to hold for objects which are so widely separated in either depth (y and D) or lateral position (x and L) that stereopsis between them is no longer an effective factor. The values of x so far considered are those between zero and L, i.e., the abscissa values of A and B in Fig. 5. If x were negative, Object C would be located to the left of Object A in Fig. 5. It would be expected from the binocular hypothesis that, in this case, some error would occur in the perception of the relative depth position of C with respect to B, and this error would be greater than that of C with respect to A. Equation 2, in this case, could be used to predict the direction of the error in the perception of C with respect to B, but it is not known whether it would successfully predict the magnitude of the error. A case similar in principle occurs when the x value of C is greater than L (the abscissa value of B). Equation 2 is expected to apply only when the value of x lies between zero and L (see Fig. 5).

Figure 5 illustrates the case in which a horizontal difference occurs in the visual direction of Objects A, B, and C. Vertical differences might also be involved. The binocular hypothesis has been stated so as to apply to differences in visual direction, regardless of the direction of these differences. Experimental evidence relating to the binocular hypothesis has been presented, however, only for the case in which the binocular illusion consisted of horizontally separated objects. Consequently, Equation 2 should be considered as applying only tentatively to situations in which vertical components of directional separation are involved in the illusion. The expectation, however, is that in such cases, Equation 2 would remain unchanged but the plane defined by the XY axes should be rotated so as to continue to contain the line AB and the line A′B′. If Point C is not in the resulting XY plane, the x value of C would probably be determined by orthogonally projecting C on this plane. The correctness of this method of determining x is suggested by the similarity of the binocular results obtained when the square was located either directly above or three inches above the large end of the rotating trapezoidal window (Positions A and B of Fig. 4). With the window illusion occurring in the horizontal plane, vertical displacement of the square did not seem to affect the apparent path of movement of the square.

From the Monocular Hypothesis

In the notation associated with Fig. 5, tentatively the monocular results of Table 1 can be described by the equation $y' = m'x$ (Equation 3). Equation 1 and Equation 3 are identical even though they involve different hypotheses. This is reflected in the similarities in the binocular and monocular results of Table 1. Equation 1 is the special case of Equation 2 which occurs when m and y are zero. Equation 3, however, is a general equation which does not involve m or y. The presence or absence of the physical terms and m and y constitutes the difference between the general equation involving the binocular hypothesis (Equation 2) and that involving the monocular hypothesis (Equation 3).

The verbal reports of the apparent depth between the two playing cards in the first experiment of this study indicate that the value of m' was larger with monocular than with binocular observation. Some of the dissimilarities between the monocular and binocular results of Table 1 can be explained by this difference in m'. For example, for Position 2 (see Fig. 2), as expected from Equations 1 and 3, the reported depth between the disc and Card A was greater when monocular rather than binocular observation was used.

Referring to Fig. 5, it is not known whether Equation 3 will apply when x is considerably negative or considerably greater than L. As with Equation 2, data relating to Equation 3 have been gathered only for the case in which A and B were horizontally separated without a large component of vertical separation. Additional cases need to be investigated.

SUMMARY

An experiment was conducted to illustrate and to extend the evidence relating to two previously formulated hypotheses involving visual direction. One of the hypotheses, called the binocular hypothesis, states that *the stereopsis between a test object and that object which has the smallest difference in visual direction from the test object is the most effective stereopsis in determining the apparent distance position of the test object with respect to each of the other objects in the field of view.* The other hypothesis, called for convenience the monocular hypothesis, states that *there is a tendency to see objects as located at the same distance from the observer, with the strength of this tendency being inversely related to the difference between the lateral visual directions of the two objects.*

To test the applicability of the hypotheses to fairly complex visual situations, the two hypotheses were used in predicting the apparent path of movement of a small object attached to the Ames rotating trapezoidal window. An experiment was reported in which the trapezoidal window was viewed either monocularly or binocularly, with four conditions of attaching the small object to the window. The differences between the results from the various conditions confirm a group of predictions made from the two hypotheses.

An equation was developed from the binocular hypothesis to predict the apparent relative depth position of a binocular object as a function of its physical relative position and the physical and apparent relative positions of other objects in the field. Also, an equation involving the monocular hypothesis was suggested.

REFERENCES

1. AMES, A., JR. Visual perception and the rotating trapezoidal window. *Psychol. Monogr.*, 1951, **65**, No. 7 (Whole No. 324).
2. GOGEL, W. C. Perception of the relative distance position of objects as a function of other objects in the field. *J. exp. Psychol.*, 1954, **47**, 335–342.
3. GOGEL, W. C. The tendency to see objects as equidistant and its inverse relation to lateral separation. *Psychol. Monogr.*, 1956, **70**, No. 4 (Whole No. 411).
4. KILPATRICK, F. P. Assumptions and perception: three experiments. In F. P. Kilpatrick (Ed.) *Human behavior from the transactional point of view*. Hanover, N. H.: Institute for Associated Research, 1952.

(*Accepted for publication November 3, 1956*)

Bureau Ednl. Psy. Research
Dated ....
No. ....

Vol. 70, No. 12 Whole No. 419, 1956

Psychological Monographs: General and Applied

Transfer and Retroaction Effects of "Association Reversal" and "Familiarization" Training in Trouble Shooting

Lawrence M. Stolurow, Thomas F. Hodgson, and John Silva[1]

University of Illinois

Industrial and military populations of employees and trainees contain several types of subgroups which differ in (presumably) psychologically significant ways. These differences can be utilized as independent variables in research directed towards discovery of *(a)* practical procedures of efficient organization of training programs in relation to work experience, and *(b)* laws describing the psychological relationships which underlie these procedures.

A general area in which these fortunate conditions prevail is that of trouble shooting. It happens, for example, that certain groups of Air Force aircraft and engine mechanics are taught about engine malfunction primarily in terms of *causes,* with which they learn to associate *symptoms.* Others begin with symptoms, and are taught to infer the possible causes. Later each type of group will be subjected to training or work experience which involves a reversal of the original sequence ("cause-symptom" or "symptom-cause"). This set of conditions can be designated *association reversal.*

It also happens that certain groups of mechanics learn to identify defective components and the names of these defects. Later on they may attend a technical school and learn to associate these names with symptoms of malfunction. Depending upon the way in which the diagnostic material is taught, i.e., as symptom-cause associations or cause-symptom, the earlier *familiarization training* received on the job could be designated stimulus training or response training.

The present study proposes to take advantage of these "natural" experimental conditions by determining their effects upon performance on specially designed "miniature diagnostic problems." Two types of problems were used to determine proficiency in engine diagnosis. They corresponded to the two types of association sequences in the training and work experience of mechanics referred to above, viz., *(a)* given the symptoms, the airman is to designate the cause(s), *(b)* given a cause he is to designate the symptoms. The nature of these problems, then, constitutes another independent variable.

If we make the assumption that learning associations in one direction, such as symptom-cause (S-C), is different from learning them in the reverse direction,

---

[1] The assistance of Mr. Gregory and Mr. K. Farson and of Mr. Binstock and the instructors who gave generously of their time to edit the problems used in this study is gratefully acknowledged. Mr. Tepper similarly contributed to this study.

This research was supported in part by the United States Air Force under Contract No. AF 33 (038)-25726, monitored by the Air Force Personnel and Training Research Center. Permission is granted for reproduction, translation, publication, use, and disposal in whole and in part by or for the United States Government.

Bureau Ednl. Research
DAVIS
Dated
No

then the learning of each constitutes a separate task and the paradigms of transfer and retroaction can be assumed to provide appropriate descriptions of the conditions that determine their interaction effects. From this point of view *association reversal* becomes a special case of these phenomena, depending upon the order in which the habits are learned in relation to the one that is tested for recall. A factor which may have significant psychological implications in relating both association reversal and familiarization training of mechanics to conventional studies of transfer and retroaction is that the associations of symptom and cause and of cause and symptom are divergent. In each case the cue is to be associated with several responses and when learned the responses presumably form a *habit family hierarchy*.

The present paper is a report of interrelated studies of both practical and theoretical problems, each of which will be described along with the experimental method used in its investigation and the results. Three variables were considered in the experimental designs: *(a)* the type of association practiced, *(b)* the sequence in which the different types of association were practiced, and *(c)* the relative amount of opportunity to practice. A sequence in which a mechanic practiced one task and then another and finally was tested for his recall of the first task was considered a *retroaction* paradigm. On the other hand, a sequence in which a mechanic practiced one task and then another and was finally tested for his recall of the second task was considered a *transfer* paradigm. It was possible to select, from a large number of examined airmen, subgroups which met the requirements of these experimental paradigms. The performance of these subgroups on the diagnostic problems was compared both to determine the effects of these variables in terms of the direction and extent of transfer and retroaction resulting from association reversal and familiarization training.

PROBLEM

The appropriate sequencing of opportunities to practice the various components of technical jobs to provide maximum transfer and minimum interference is a persistent and fundamental problem. A necessary type of research in relation to this problem is the determination of the interaction effects which different opportunities to practice task components have upon one another.

Methodological Requirements

One reason why this type of research has not been conducted on a more extensive scale is the severe requirements that must be met in order for results to be clearly interpretable. One of these requirements is the appropriate characterization of the tasks which are separately practiced. Secondly, it is necessary to identify the sequences in which the appropriate opportunities to practice have occurred in the training and work histories of sizable groups of individuals for whom other potentially influencing factors are controlled. In addition, it is necessary to determine that the distribution of talent, as measured by an aptitude test, for example, is the same for the groups being compared. Having satisfied these requirements, it is possible to compare the performance of groups that are equated on a number of variables and different in a specified way on another variable.

In accomplishing these comparisons it is important to sample both problems and subjects rather broadly. In trouble shooting Air Force powerplants, both line and dock experience are subject to a multiplicity of factors, and variations from individual to individual are to be expected. Depending upon the mechanic's job assignment and the length of time he works at this type of task, he will receive different amounts of practice on the various defects that might occur in a particular device. The frequency of occurrence of every defect, the length of time the man works on a "shred-out" job, and the types of shred-out jobs he is assigned are some of the factors that determine the amount and type of opportunity a mechanic has to practice the associations critical to the present research. The net result is that these experiences provide a variety of different degrees and types of practice with the possible diagnostic problems that could arise with a complex device such as a reciprocating engine.

Types of Associations

Divergent associations or multiple linkages. Two features of symptom-cause and cause-symptom associations appear to have important psychological implications. A symptom of malfunction, in many if not most complex systems or devices, is associated with more than one possible defect or cause. Also a defect may be indicated by more than one symptom of malfunction. These types of cue-response relationships have been designated *divergent* association. These multiple linkages between terms appear to be an important set of conditions with respect to learning, transfer, and retention of much technical information such as that used in diagnosing equipment failures. As compared with the learning, transfer, and retention of paired-associates types of connections, multiple linkages or divergent associations can be described as: *(a)* one to many; *(b)* many to one, and *(c)* many to many. In these terms, traditional paired associates are one-to-one linkages.

Sharing and overlapping linkages. Another presumably important feature of symptom-cause and cause-symptom associations is that a symptom (e.g., S_1) may be associated with a defect (D_1) that also produces one, or more, other symptoms of malfunctions (e.g. S_2); in addition, the reverse frequently is true. Thus the two symptoms (S_1 and S_2) also may be linked to a second defect (D_2). Many of the associations involved in learning diagnostic information, therefore, have *overlapping* linkages, i.e., terms in one set—symptoms—are linked to two or more terms in a second set—defects—and correspondingly, each term in the second set is linked to two or more of the same terms in the first set. In other words the pattern of linkages in both directions can be described as divergent connections (16).

If a single stimulus term is to be linked with two or more responses (one-to-many linkage), the relationship can be described as response sharing; conversely when two or more S terms are linked to one R term, they constitute many-to-one associations. The relationship will be designated stimulus sharing; similarly, many-to-many associations can be described as stimulus and response sharing.

Types of Practice

Three types of situations, with respect to the opportunity to practice particular types of relevant associations, figure in the comparisons made in the present study. Each type of association can be identified in terms of the setting permitting its practice, i.e., technical school training, line experience, and dock experience.

Technical school. The instruction in the reciprocating engine course which figured in this study is probably typical of that found in many, if not most, courses dealing with the location of defective components in complex systems. That they would have this feature in common is particularly likely if they share the educational philosophy of this course which is to teach the student to "understand" the equipment rather than a set of "cookbook" procedures to be used in the solution of trouble-shooting problems. When malfunction states are discussed with an emphasis upon teaching for understanding, it is rather natural to use a so-called "logical" order of presentation. However, when alternative and equally logical accounts exist, a decision is required. In the case of diagnostic information there are two choices. One is an account of events in the temporal order of their occurrence which might be designated the *physical order*. The second is an account of events in the order of discovery in trouble shooting which might be designated the *diagnostic order*.

Using the physical order, a typical classroom discussion might be started by assuming that a particular defect exists in a system of the equipment. Next, the implications of this hypothetical defect with respect to the operation of the device would be traced. Each assumed defect serves to make possible a discussion of the operation of the equipment when the defect is present and a review of the normal operation of the system. While this procedure has much to recommend it, its psychological implications previously have not been submitted to analysis or experimental study.

In the advanced technical school which figured in this study, the emphasis was upon teaching symptoms produced by defects or causes, i.e., C-S associations, and less stress was given to the reverse direction of association, i.e., S-C associations. This emphasis can be represented as C-S $>$ S-C.

Line experience. Field observations and interviews with mechanics and instructors who had job experience[2] permitted the characterization of line and dock work in terms of the predominantly, if not exclusively practiced cue-response relationships which these jobs permit. Line experience may be characterized as predominantly S-C learning. In performing this type of work the mechanic begins to solve a diagnostic problem knowing the indications of malfunction and is required to determine the defect producing the particular symptoms he observed.

Dock experience. Dock experience may be described as C $>$ S learning since it provides the mechanic with an opportunity to learn C terms or the labels associated with causes (or defects) and with little if any opportunity to learn S terms (or the labels of symptoms). In working on the dock, the mechanic performs preventive maintenance. He takes apart engine components and inspects them to determine whether there are defects present. In doing this it is likely that he would learn the names of the defects he observed; however, he would not be required to practice the association of symptoms with these defects or vice versa.

Sequences of Air Force Training and Experience in Trouble Shooting

At the time the present study was conducted it was possible for a variety of combinations of training and work experience to occur; however, a sizable but varying proportion of airmen in the advanced reciprocating engine course would be "pipeline" students. These men were sent directly to the advanced course from a basic school without intervening job experience. Others in the same class in technical school may have acquired different amounts and types of related work experience before attending the advanced course. Still others would have completed the advanced course as pipeline students and then acquired job experience.

The *phenotypic* variables (7, p. 11) associated with the problems studied are *(a)* training and work sequence, *(b)* amount of work experience, and *(c)* type of work experience. In addition to the determination of the effects of these variables as such on the recall of C-S and S-C associations, appropriate comparisons also can be made which take into consideration the sequence in which the various types of associations were practiced, i.e., *genotypic* variables. When the sequence is taken into consideration and related to the type of diagnostic problem used to measure retention, the conditions are met for either a transfer or retroaction design involving association reversal or familiarization training.

---

[2] These are described and summarized in unpublished technical reports transmitted to AF Personnel and Training Research Center in 1952 and 1953.

Transfer effects. The performance of pipeline students on problems which require the recall of causes of engine malfunctions, given the symptoms (S-C or symptom-cause problem), and the recall of symptoms that could result from stated causes (C-S or cause-symptom problems) will reflect the level of proficiency generated by training in the advanced reciprocating engine course. The scores of this group therefore could be used as a control measure for the purpose of comparing groups with different work experience, all of whom attended the advanced reciprocating engine course. In this way it is possible to determine the relative amount and direction of delayed transfer effects resulting from the various conditions of work.

The notion of delayed transfer effects as used here refers to the fact that the effects of transfer are being measured after the trainee has had considerable opportunity to practice the second task on which the test of transfer is based. This condition is different from the determination of transfer after various intervals between learning Task 1 and Task 2. Data from laboratory studies indicate that the intertask interval has very little effect on the amount of transfer that results from training on the first task (2, 3, 8, 12, 15). This is a fortunate circumstance, since in this study it was not possible to control the intertask interval.

Transfer effects attributable to *association reversal* could be determined under several conditions. For example, both brief and longer *line* experience can serve as the first task, in which case the scores from the C-S problems could be used as the dependent variable. Similarly the transfer effect of advanced technical school training can be determined from the scores on S-C problems for the mechanics who received their job experience (S-C associations) after attending technical school.

There have been several studies of transfer after different amounts of practice on Task 1 (8, 12). The typical finding is that increasing amounts of practice on the first task results in increased amounts of positive transfer effect. In fact, even if a small amount of practice on Task 1 resulted in negative transfer effects, large amounts of practice have produced positive transfer effects (8). This finding would lead to the expectation that the same might hold in the case of association reversal.

The transfer effects of familiarization training also can be determined for those who had dock rather than line experience. The scores on the C-S problems provide an index of the effects of *stimulus training* and those based upon S-C scores represent the effects of *response training.* In both cases the conditions depart from those conventionally studied since *(a)* divergent overlapping associations are involved rather than simple S-R connections, and *(b)* training was probably selective.

Results from studies of predifferentiation training (1, 6, 9, 10, 14) would lead to the expectation that stimulus training might produce positive transfer effects if the cue stimuli are confusable; however, if they are not confusable, then it would appear that the results of "associative inhibition" studies (8) would apply and the learning of C-S associations after learning the names of defects would be slower than if the student had not had the dock experience. Since there are no data upon which to base a decision about the confusability of these terms, it is not possible to state in advance which one of these notions would apply in the present research.

Since appropriate data are not available, the effects of response training can not be predicted. In addition to the factors of divergence and overlap, among the S-C and C-S associations, the response training may have been selective, i.e., disproportionally distributed, as a consequence of: *(a)* the differences in the probability with which the various defects appeared, *(b)* the job assignments of the men, and *(c)* the emphases in school.

Retroaction effects. The conditions obtaining for various groups of Air Force mechanics also permit the determination of the retroaction effects of association reversal under some specifiable conditions. Before attending the advanced reciprocating engine course the mechanics with line experience learn S-C associations, and then in school they learn C-S associations. The recall scores on S-C problems represent the retroaction effects of association reversal when compared with those of the pipeline students. Also for groups differing in the amount of work experience, it is possible to determine the extent of retroaction when the opportunity to practice the earlier learned task varied in amount.

The traditional concern of retroaction studies has been with the conditions that are related to interference effects produced by the learning of an interpolated task. This emphasis is probably explained by the historical interest in forgetting. The most commonly accepted explanation of losses in ability to perform a task is that it is due to the interpolated learning of an interfering task (8, 12). Examples of conditions found to be significant in this respect are the interrelationships between the stimuli in the two tasks, between the responses, and between S-R associations in each case. With one exception, association reversal is one type of intertask relationship previously overlooked (13). Knowledge of the retroaction effects of this variable under the conditions obtaining for various groups and trainees therefore would be of both practical and theoretical interest.

METHOD

Toops's (17) addend coding procedure was used to identify the individuals comprising the five experimental groups that were compared. This was made possible by including a biographical questionnaire in the problem booklets.

Taxonomic tables were prepared for the two groups of airmen used in this study. In both tables four classes of experience were used, each of which was represented by a number of variations. In the taxonomic table for the airmen in technical school the experience classes were: *(a)* basic training school, *(b)* type of engine experience, *(c)* amount of experience in months, and *(d)* type of job preceding school. For the airmen working as mechanics the first three classes were the same; the fourth was present job assignment.

Experimental Paradigms

Two different but interrelated sets of comparisons figured in the present study, i.e., the "phenotypic" and "genotypic." One set was designed to obtain information related to the practical problems of organizing technical school training and job experience; the other set was used to obtain information related to problems of transfer of training and retroaction under conditions of association reversal and familiarization training.

Sequences of technical school and job experience. This first paradigm was used to determine the relative effects of different sequences of work experience and technical school training. Two different sequences were compared with the performance of pipeline students on the same problems. The comparisons that were made can be summarized as follows:

Group		Experience and Training		Recall Tests
Code	Description	1st task	2nd task	
CG	Controls (pipeline students)	(None)	Advanced training C-S>S-C	(a) S-C (b) C-S
EA	Experimental condition *A*	Brief line experience S-C	Advanced training C-A>S-C	(a) S-C (b) C-S
EB	Experimental condition *B*	Advanced training C-S>S-C	Brief line experience S-C	(a) S-C (b) C-S

Amount of work experience. The paradigm used to determine the effects of amount of work experience preceding technical school training was:

Group		Experience and Training		Recall Tests
Code	Description	1st task	2nd task	
CG	Controls (pipeline students)	(None)	Advanced training C-S>S-C	(a) S-C (b) C-S
EA	Experimental condition *A*	Brief line experience S-C	Advanced training C-S>S-C	(a) S-C (b) C-S
EC	Experimental condition *C*	Longer line experience S-C	Advanced training C-S>S-C	(a) S-C (b) C-S

Type of work experience. The paradigm used to determine the effects of differences in the type of work experience was:

Group		Experience and Training		Recall Tests
Code	Description	1st task	2nd task	
CG	Controls (pipeline students)	(None)	Advanced training C-S>S-C	(a) S-C (b) C-S
EA	Experimental condition *A*	Brief line experience S-C	Advanced training C-S>S-C	(a) S-C (b) C-S
ED	Experimental condition *D*	Brief dock experience C>S	Advanced training C-S>S-C	(a) S-C (b) C-S

Transfer effects. Examination of the first paradigm indicates that the C-S scores of the EA group on the recall tests represent the delayed effects of transfer from S-C training, *association reversal.* These mechanics had learned S-C associations during their brief period of work on the line which preceded their training in advanced technical school. This latter training provided them with the opportunity to learn C-S associations. Therefore if the C-S recall scores of the EA group are compared with those of the pipeline students (CG), it is possible to determine the amount and direction of the transfer effect of their work experience. If the C-S scores of the EA group were to be smaller than those of the CG it would indicate that the line training which the EA group had received before attending technical school had a negative transfer effect; the reverse would indicate positive transfer.

The second paradigm deals with the effects of different amounts of similar experience prior to taking the advanced technical school course. The C-S recall scores are an index of the delayed transfer effects of line experience. This latter experience permitted the mechanics to practice S-C associations. This was the Task 1 condition for both EA and EC groups, but in the former case the *amount* of opportunity to practice the S-C associations was less than in the latter.

In a similar way the transfer effects of *familiarization* training can be determined under two different conditions. These are provided by the third paradigm through the comparisons of the ED group with the CG on the C-S problems. This comparison indicates the effects of stimulus training, and on the C-S problems the comparison of these groups indicates the effects of response training.

Retroaction effects. In the first paradigm the S-C scores of the EA group in relation to those of CG represent the retroaction effects of advanced technical school training. Similarly, the C-S scores of the EB group in relation to those of CG indicate the amount and direction of the retroaction effects of brief line experience.

In the second paradigm the S-C scores of the EC group in relation to those of CG represent the retroaction effects of advanced technical school training after a longer opportunity to practice S-C associations than that available to the EB group in the first paradigm.

Subjects

A total of 297 airmen were examined while at technical school and an additional 80 while working as mechanics (at two different air bases). The five groups which would permit the comparisons indicated in the paradigms are presented in Table 1.

The technical school provided sufficiently large numbers of airmen in CG (ulstrith 284),[3] EA group (ulstrith 845), EC group (ulstrith 900), and ED group (ulstrith 1945); and the two air bases provided samples, in the EB group (ulstrith 30), sufficiently large to permit comparisons with the others as indicated. Of the 377 airmen who were examined however, only 178 met the requirements of this study. These were: *(a)* appropriate biographical history to permit classification into one of five ulstriths, and *(b)* completion of all the problems.

Descriptions of only the critical biographical data are presented in Tables 1 and 2. To save space, complete descriptions of the biographical variables used in the taxonomic tables are not presented in this report. Table 1 presents the information that is important to the present study.

The clerical and mechanical aptitude indices from the Airman Classification Test Battery were secured for all airmen.

---

[3] Ulstrith is a term coined by Toops (17). An ulstrith code number is the sum of addends assigned to the variations in the classes of biographical data considered to be relevant and important to the study. "Addends are numbers which when added yield a code number uniquely identifying the ultimate breakdown society to which a given person belongs" (Toops, 17, p. 41). Toops indicates the rules to be used in assigning these numbers; however, the needs of a particular study determine the events and variations of each which are included in the taxonomic table.

TABLE 1

Composition of Airmen Samples in Terms of Selected Biographical Information*

Classes of critical biographical data	Ulstriths				
	At end of training				On the job
	CG 284	EA 845	EC 900	ED 1945	EB 30
Job experience					
Student	x				
Line mechanic		x	x		x
Dock mechanic				x	
Amount of experience (in months)					
None	x				
0 to 8.9		x		x	x
9 to 17.9			x		
Engine experience					
None	x				
R-4360 only		x	x	x	x
Sequence of school and job					
School before job experience	x				x
School following job experience		x	x	x	

* A total of 297 airmen were examined while in training and 80 while working as mechanics at two SAC bases. Of this total, 178 met the requirements of this study. The numbers given each problem-booklet form were as follows:

Form A (94)	31	25	10	17	11
Form B (84)	32	14	10	19	9

Problems—Recall Test Materials

The problems used were adapted from those employed in the evaluation of trainees in the reciprocating engine course given in technical school. They are particularly modeled after the type of problem used to examine trainees in the trouble-shooting phase of the Advanced Reciprocating Engine Course. The problems were systematically edited and revised. This was a cooperative enterprise involving personnel associated with the course and the authors.

Description of problems. Twenty-one problems were of the type "List the symptoms caused by ————." These are referred to as the cause-to-symptom (C-S) problems. The trainee is required to associate a list of symptoms with a given cause. The remaining 29 were of the type "List the possible causes of the following symptom: ————." These were described as symptom-to-cause (S-C) problems.

Problem booklets. Since it was apparent that the presentation of 50 problems of these types to any one airman would overburden him, particularly since each list of symptoms and causes may be fairly long, it was decided to divide the total set of problems into two equivalent sets. Accordingly, two sets of the problems were assembled into booklets each of which contained 25 different problems. Form A, Section I contained 10 problems of the C-S type, and Section II contained 15 S-C; Form B contained 11 and 14 problems of each type, respectively. The two forms were made as comparable as possible.

Each form had the same set of initial pages. Each airman filled in the blanks on these pages with his biographical data. All booklets also contained the same sample problems and instructions to the airman. The two forms differed in that they did not contain common problems.

The problems were all of the write-in type. Either a symptom or a cause of a malfunction was stated and the airman was obliged to recall and write down as many relevant responses as he could think of.

Classification of Responses and Scoring

Since a "consensus" of expected right answers was available for every problem, it was possible to classify most of the correct responses under each of these. Where responses did not fit under a rubric of the scoring key, they were classified as either ambiguous or uncertain. Every problem was separately tabulated.

A classification system for responses to C-S problems. Four general categories of response were: right, wrong, ambiguous, and uncertain. Under each of these rubrics there were response types under which individual "tokens" were tallied. The consensus of expected right answers serves as the types for the general category. Those for the wrong responses were generated by the data, as were the ones included under the other two general categories. An ambiguous response was one which we were not able to decide upon with respect to its correctness or incorrectness because the answer was incomplete. The uncertain responses were those which were apparently complete but as stated were assumed to be highly improbable. In most cases a response was classified as uncertain when the magnitude or degree of the defect or cause would be the critical determinant of the presence or absence of the symptoms and the trainee failed to say anything about this.

The various response types under each of these general classes were organized under each of three subclasses: *(a)* symptoms, *(b)* malfunction conditions or states, and *(c)* defective physical units or causes. Only symptoms and malfunction conditions met the requirements of a correct response to the C-S problems (Sections I in Forms A and B).

A classification system for responses to S-C problems. The processing and analysis of the responses to questions in Section II were conducted in the same general manner as described for Section I. The two major categories were right and wrong, and the latter was subdivided into: *(a)* wrong causes, *(b)* ambiguous causes, and *(c)* symptoms. No attempt was made to subdivide these problems according to the powerplant systems since a given symptom may have been the result of defects in any one of several different systems. There were 15 problems of this type in Form A and 14 in Form B.[4]

Scoring of C-S problems. The responses under the categories ambiguous and uncertain were given a weight of zero. For Section I arbitrary values of plus one were assigned to every right response and minus one to every wrong response. The score for every airman was an algebraic total and therefore could be zero, or a minus or plus value depending upon the relative frequency of correct and incorrect responses. A constant positive integer—10—was added to all scores to make them all positive and larger than zero.

Both Forms A and B were scored in the same way; however, Form A contained 10 and Form B 11 C-S problems. The scores for two different subsets of problems were determined depending upon whether they were *fuel system* problems or problems from the other engine systems (general). This separate analysis was made in anticipation of differences in sensitivity to the effects of the experimental conditions. The fuel system happened to be the only one with a sufficient number of problems to warrant separate scoring. This was simply the joint result of the problems that were available to work with and the editorial processing.

---

[4] One problem, Number 5 of Form A, Section II, was not scored.

Both Forms A and B were scored using the simple binary weights (±1). An algebraic total score was computed for every airman and these scores were summed for each group (ulstrith).

Procedure

Arrangements were made with the instructional supervisors of the Reciprocating Engine Division in technical school to administer the problems during class hours to all trainees enrolled in the Engine Course. The last (fourth) phase of the course was chosen as the one most suitable for the experiment. Ten separate classes were examined on the sixth day of the last phase.

The instructors were asked to introduce the problems as though they were a regular part of the course requirement. It was pointed out to the instructors that a student might ask if his score would count in his final grade. In anticipation of this the instructors were told to answer in an equivocal manner, such as "All your work is important here." This answer at the very least did not specifically identify the problems as an experimental device and probably insured an adequate level of motivation.

Instructors passed copies of the instruction sheets to their students. They then read aloud the instructions while the students followed by reading their own copies. At the appropriate place in the directions the instructor distributed the Problem Booklets to the students.

The Problem Booklets previously had been arranged in consecutive numerical sequence which in effect resulted in the distribution of Forms A and B to alternate members in the classroom. This procedure allowed approximately equal distribution of Forms A and B at any one testing, and also helped to reduce the amount of cooperative work among the trainees. Also it was assumed that this procedure would result in approximately randomized distribution of problems over the available talent. There was no convenient way of determining in advance of the experiment the actual ability levels of the individuals in every class. A great deal of investment in time, examining personnel records, and tracing individuals through class rosters would have been necessary to make this possible. Reliance was placed, therefore, upon the probable chance distribution of talent over the two forms of problems rather than upon any selection procedure involving the identification of individuals. A check was made of the homogeneity of the groups by later examining their aptitude scores.

First the trainees filled out the personal history items. Then the instructor told them that they had 60 minutes to complete the 25 problems in their booklet. Following this he told them to begin. At the end of 60 minutes, trainees were told to close their booklets and place them on their desks where they would be collected by the instructor.

During each session the research project representative sat at the back of the room and observed the procedure used. He was prepared to note disturbances and exceptional conditions on a specially prepared check sheet.

While the testing sessions were being conducted in technical school, arrangements were made to have the problem booklets administered to a small sample of airmen who were working as mechanics. As far as possible, the procedure used was the same as that employed in technical school.

RESULTS

One set of groups of subjects (ulstriths 284, 845, and 30) made possible the determination of the effects of *order of experience* with respect to school training and work experience as indicated in the first paradigm. Another made possible the determination of the effect of different *amounts of experience* before attending technical school (ulstriths 284, 845, and 900). A third set made possible the determination of the effects of different *types of experience* (ulstriths 284, 845, and 1945). An analysis of the pertinent constant and variable factors in the experience of these five groups is indicated in Table 2.

Subjects in ulstrith 284, the pipeline students, were designated as CG in the paradigm analysis. They are included in

TABLE 2

DESCRIPTION OF PARAMETERS DETERMINING COMPARISONS OF RESPONDENT ULSTRITHS

Groups Compared*	Constants**			Variables		
	School sequence	Type of experience	Amount of experience	School sequence	Type of experience	Amount of experience
(1) 284, 845, 30		x	x	x		
(2) 284, 845, 900	x	x				x
(3) 284, 845, 1945	x		x		x	

* Ulstrith numbers. CG is ulstrith 284, EA is 845, EB is 30, EC is 900, and ED is 1945.

** An important feature of the use of Toops's method is the control permitted by identifying potential variables which can be held constant while a single "molar" variable is studied. The parameters identified as constants in this table were constant for the groups compared in a particular analysis. The addends procedure resulted in five ulstriths for which the individuals had a large number of common biographical events.

each comparison to serve as a standard or control group.

Aptitude indices. The mechanical and clerical aptitude indices of these groups were tabulated as a control measure. When these aptitude indices were examined for statistical significance using the distribution-free test described by Mood (11, p. 398), neither of the analyses —groups tested with Forms A and B— resulted in a significant X^2 (minimum confidence level of five per cent). This means that all groups were equivalent in talent and therefore that a difference in their S-C and C-S scores is not attributable to aptitude.

Analysis of Variance of Ulstrith Scores

Three sets of variance analyses were computed. They corresponded to the paradigms previously described. In each set there were three $3 \times X$ designs in which the first factor consisted of the three *groups* (ulstriths) in the paradigm and the second was a set of *problems.* There were six analyses for every paradigm—three types of problems in each of two separate sets or booklet forms. Since an airman responded to all of the problems in a booklet and every booklet contained three sets, the second factor in the analyses of variance represented repeated measurements on the same subjects. Therefore the problem mean square was evaluated in terms of the pooled $Ss \times$ problems mean square (4, pp. 289-296). Section I of both Forms A and B contained the C-S problems, and these constituted one analysis for every paradigm. Section II contained two different types of problems: *(a)* S-C general systems except fuel, and *(b)* S-C fuel system. Each of these three sets of problems in each form constituted a separate analysis of variance for a particular set of three subject groups.

The results of these analyses are presented in Table 3 summarized under the three phenotypic variables: *(a)* school-work sequence, *(b)* amount of line experience, and *(c)* type of experience. The two principal variance components are: *(a)* ulstrith differences, and *(b)* problem differences. The individual problem means are presented in Tables 4 and 5.

Problem differences. The most consistent result from the variance analyses was the significant difference between problems. (See Tables 3, 4, and 5.) In every analysis, the between-problems variance was sufficiently large to produce significant F ratios (beyond the .01 level).

The methodological implication of this result is clear. It indicates the hazard

TABLE 3

SUMMARY OF RESULTS FROM VARIANCE ANALYSIS OF RECALL SCORES ON PROBLEMS

Paradigm	Variable	Groups (ulstriths)	Type of Problem	Booklet	System	F Ratios: df b/w	F Ratios: Between ulstriths	F Ratios: df b/w	F Ratios: Between problems***
1	School-job sequence	CG, EA, EB (284, 845, 30)	S-to-C	A	General	2/64	13.22***	13/832	60.61
			S-to-C	B	General	2/52	4.20**	13/676	24.25
			C-to-S	A	General	2/62	2.88*	5/310	22.45
			C-to-S	B	General	2/52	3.86**	5/260	34.54
			C-to-S	A	Fuel	2/63	5.35*	3/189	69.65
			C-to-S	B	Fuel	2/52	.69	4/208	9.77
2	Amount of line experience	CG, EA, EC (284, 845, 900)	S-to-C	A	General	2/63	.55	13/819	60.11
			S-to-C	B	General	2/53	.56	13/689	26.86
			C-to-S	A	General	2/62	.29	5/310	22.48
			C-to-S	B	General	2/52	.21	5/260	30.36
			C-to-S	A	Fuel	2/63	4.85**	3/189	69.00
			C-to-S	B	Fuel	2/55	.80	4/212	13.02
3	Type of experience	CG, EA, ED (284, 845, 1945)	S-to-C	A	General	2/70	.27	13/910	65.68
			S-to-C	B	General	2/62	.68	13/806	31.51
			C-to-S	A	General	2/68	.59	5/340	27.79
			C-to-S	B	General	2/62	1.41	5/310	38.77
			C-to-S	A	Fuel	2/69	2.18	3/207	73.68
			C-to-S	B	Fuel	2/63	.26	4/244	12.06

* Probably significant, *F* lies between the 10 and 5 per cent levels for the appropriate degrees of freedom.
** Significant, *F* lies between the 5 and 1 per cent levels for the appropriate degrees of freedom.
*** Very significant, *F* lies beyond the 1 per cent level for the appropriate degrees of freedom.

of generalizing results from a small set of problems to any other set. The R-4360 powerplant is sufficiently complex to make it very unlikely for any one to achieve high *general* level of proficiency in trouble shooting. Mechanics probably learn some problems, say the more frequent ones, very well, and others a little or not at all, depending upon the opportunities provided by their training and job assignments.

School-job sequence. Five out of six *F* ratios were significant in the analyses of variance computed to determine the effects of the school job sequence. (See Table 3.)

TABLE 4

MEANS FOR SYMPTOM-TO-CAUSE PROBLEM*

Problem No.	Problem Form	Ulstriths** 284 CG	845 EA	900 EC	1945 ED	30 EB
1	A	14.06	13.94	14.09	13.83	10.96
2	A	13.15	13.35	14.18	13.51	10.86
3	A	12.56	12.44	13.76	12.82	10.51
4	A	19.19	17.86	18.23	18.21	14.61
5	A	14.42	14.49	14.99	14.76	11.74
6	A	14.47	14.44	15.04	13.42	11.39
7	A	12.51	12.76	13.58	13.06	10.79
8	A	13.20	13.29	13.10	13.19	12.27
9	A	12.89	13.49	13.99	12.29	11.50
10	A	12.67	12.65	12.48	11.92	11.81
11	A	9.88	10.15	9.74	9.51	10.61
12	A	13.31	13.33	13.00	13.04	11.95
13	A	13.50	13.28	14.51	12.24	11.17
14	A	9.89	10.17	10.49	10.52	10.12
Means		13.26	13.26	13.66	13.02	11.45

Problem No.	Problem Form	Ulstriths** 284 CG	845 EA	900 EC	1945 ED	30 EB
1	B	12.84	13.07	12.00	13.11	11.56
2	B	11.59	12.29	13.3	12.16	10.78
3	B	10.00	10.36	11.1	10.89	10.33
4	B	12.72	12.78	12.3	12.21	12.00
5	B	10.09	10.00	11.0	10.21	9.56
6	B	10.25	10.21	10.2	9.63	10.11
7	B	11.18	10.93	11.4	10.87	10.56
8	B	9.41	9.71	8.9	9.26	10.11
9	B	10.09	10.07	9.1	10.63	10.22
10	B	11.41	12.29	11.8	11.74	10.80
11	B	10.81	10.71	10.4	10.84	9.78
12	B	10.91	11.43	11.9	10.84	10.56
13	B	12.34	12.00	11.8	12.52	10.89
14	B	12.03	11.86	11.9	12.63	11.33
Means		11.12	11.26	11.22	11.25	10.62

* Differences were significant for both Forms A and B among ulstriths 284, 845, and 30 demonstrating the differences in school-job sequence (see Table 3).
** See Table 1 for description of groups.

Bureau Ednl. Psy. Research
DAVID ... COLLEGE
Dated
No

TABLE 5

MEANS FOR CAUSES-TO-SYMPTOM PROBLEMS

Problems		Ulstriths					Problems		Ulstriths				
Number	Form	284 C	845 EA	900 EC	1945* ED	30 EB	Number	Form	284 CG	845 EA	900 EC	1945* ED	30 EB
							Fuel system**						
3	A	10.26	11.08	11.20	10.75	6.19	3	B	11.06	11.50	12.00	11.56	11.00
4	A	10.58	9.96	12.30	11.19	5.75	4	B	11.94	12.29	13.00	11.72	11.67
5	A	13.71	13.72	15.20	14.81	7.88	5	B	11.53	11.79	12.50	12.22	11.00
6	A	13.77	13.52	14.30	14.19	7.50	6	B	10.31	10.43	9.20	9.50	10.00
							10	B	11.75	11.71	12.20	11.22	11.44
Means		12.08	12.07	13.25	12.74	6.83	Means		11.32	11.54	11.78	11.24	11.02
							General***						
1	A	11.60	12.40	12.10	12.19	9.60	1	B	11.53	11.57	11.67	12.21	9.78
2	A	9.83	8.60	9.90	9.94	10.00	2	B	9.22	9.00	9.44	9.89	8.22
7	A	12.47	10.30	12.00	13.00	12.00	7	B	10.25	10.00	9.67	10.26	9.22
8	A	11.00	9.13	11.20	11.06	10.20	8	B	10.81	10.64	10.67	10.63	10.11
9	A	10.83	8.97	11.90	11.25	10.70	9	B	11.88	11.71	11.67	12.47	11.55
10	A	10.07	8.13	10.00	10.19	9.50	11	B	9.63	9.36	9.89	9.53	9.67
Means		10.97	9.59	11.18	11.27	10.33	Means		10.55	10.38	10.50	10.83	9.76

* This group did not enter into either of the variance analyses that produced significant effects among ulstriths.
** Significant differences in amount of line experience (284, 845, 900), see Table 3.
*** Significant differences in school-job sequence (284, 845, 30), see Table 3.

Amount of experience. These ulstriths differed sufficiently to produce a significant *F* ratio for the fuel system C-S problems (Form A).

Type of work experience. No significant *F* ratios were obtained from these comparisons.

Transfer Effects

In determining the effects of genotypic variables under conditions of association reversal and familiarization training, the relative amounts of transfer and their direction were computed by obtaining the differences between the experimental groups and the controls, dividing this by the score of the controls (pipeline students—CG) and multiplied by 100. These data are presented in Tables 6 and 7.

School-job sequence. If we examine the transfer effects reflected by the C-S

TABLE 6

PERCENTAGE OF TRANSFER EFFECTS RESULTING FROM "ASSOCIATION REVERSAL"

Groups Compared with Ulstrith 284 (control group)		Training (Task 1)		Recall Test (Task 2)		Transfer Effects	
Group	Ulstrith	Type	Amount	Form	Problems	Negative	Positive
EA	845	Line	(0 to 8.9 mo.)	A	C-S (general)*	0.08	
				B	C-S (general)*		1.94
				A	C-S (fuel)	12.58	
				B	C-S (fuel)	1.61	
EB	30	Technical school		A	S-C	13.65	
				B	S-C	4.49	
EC	900	Line	(9 to 17.9 mo.)	A	C-S (general)		9.69
				B	C-S (general)		4.06
				A	C-S (fuel)		1.91
				B	C-S (fuel)	0.47	

* See text for distinction between general systems and fuel system problems.

TABLE 7

PERCENTAGE OF TRANSFER EFFECTS RESULTING FROM "FAMILIARIZATION TRAINING"

Group Compared with Ulstraith 284 (control group)		Training (Task 1)	Recall Test (Task 2)		Transfer Effects	
Group	Ulstrith		Form	Problems	Negative	Positive
ED	1945	Response training	A	S-C	15.16	
			B	S-C		1.16
		Stimulus training	A	C-S (general)*		5.46
			B	C-S (general)	0.71	
			A	C-S (fuel)		2.73
			B	C-S (fuel)		2.65

* See text for distinction between general systems and fuel system problems.

scores, we find a sizable amount of negative transfer for the EA group (ulstrith 845) on the fuel system problems (Form A). Also, there were sizable negative transfer effects obtained with group EB (ulstrith 30) as indicated by the S-C scores.

It appears that association reversal can produce sizable negative transfer effects regardless of whether the mechanic first learns S-C associations on the job and the C-S information in school, or vice versa. These results were obtained with independent groups that had had nine months or less of work experience on the line. These data (Table 6 and Fig. 1) could be interpreted as an indication that: *(a)* a small amount of job experience before advanced technical school training may interfere with school performance on C-S problems, and *(b)* more importantly, school training which emphasizes C-S associations may produce habits that interfere with the mechanic's performance of his job up through the first nine months.

Amount of experience. The data showed positive transfer effects for the more experienced line mechanics—group EC (ulstrith 900). This was found for the fuel system C-S problems (Form A) as well as for the general systems problems (Forms A and B). These mechanics had from 9 to 18 months of line experience

TABLE 8

RETROACTION EFFECTS ASSOCIATED WITH INTERPOLATED SCHOOL TRAINING AND "WORK" EXPERIENCE

Groups Compared with Controls (ulstrith 284)		Problems		Retroaction	
		Form	Type	Source	Percentage
EA	845	A	S-C	School training	0.0
		B	S-C	School training	+1.26
EC	900	A	S-C	School training	+3.02
		B	S-C	School training	+0.90
EB	30	A	C-S (general)	Work experience	−43.46
		B	C-S (fuel)	Work experience	−5.83
		A	C-S (general)	Work experience	−2.65
		B	C-S (fuel)	Work experience	−7.49

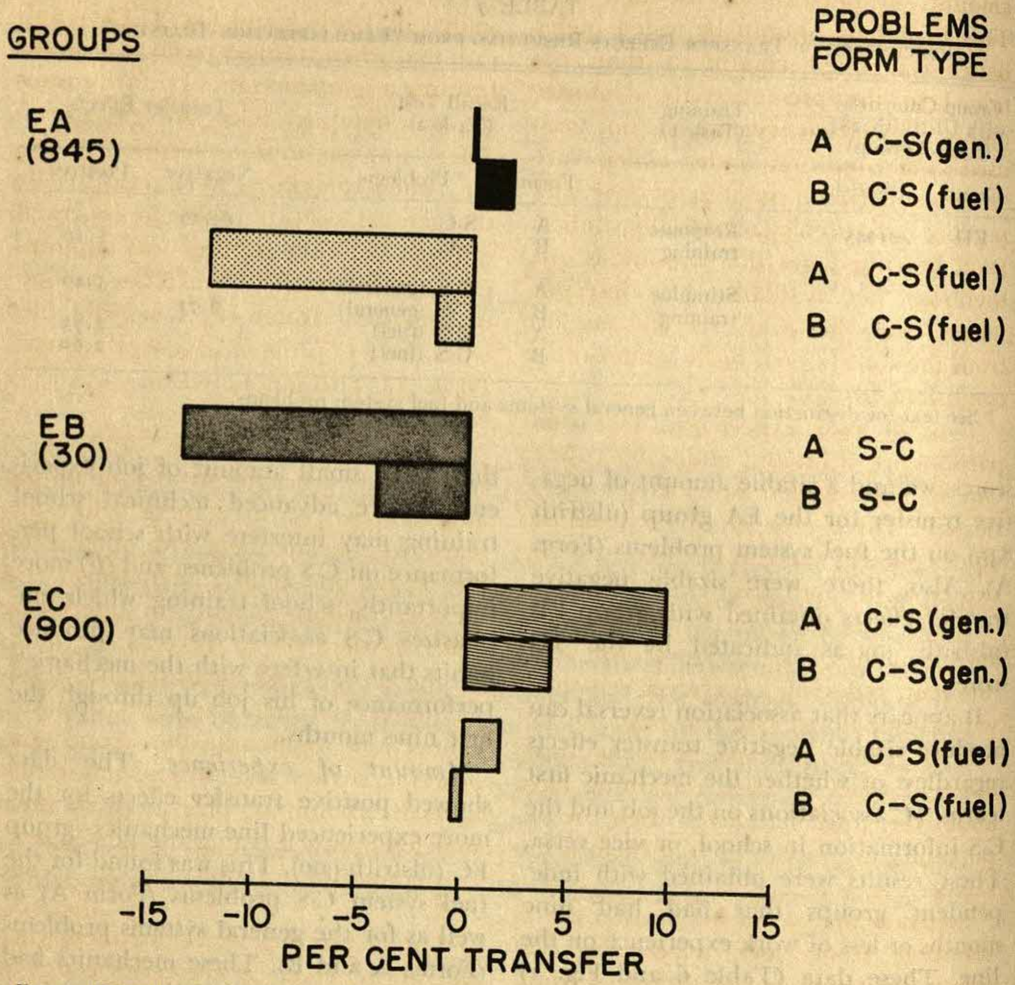


FIG. 1. Percentage of transfer effects in relation to the performance of "pipeline" students as measured by the recall scores on different types of problems. EA and EB groups had 0-8.9 months of line experience (Task 1); EC group had 9-17.9 months of line experience (Task 1). See Table 3.

before attending the advanced technical school course. These data could be interpreted to mean that after a fairly large amount of training on the first task—learning to associate terms in Set B with those in Set A—the learning of a second task may be facilitated when the second one requires the trainee to reverse the direction of association of the terms used in Task 1 but not their relationship to each other. Combining this result with that obtained with mechanics who had less training before transferring to a cue-response reversal task of the type described, it is possible to infer a testable hypothesis about the relationship that might exist between the interaction effects obtained with association reversal on the one hand and the amount of training on the first task on the other hand. It could be inferred, for example, that under association reversal with complex materials involving divergent and overlapping linkages, decreasing amounts of interference during second task learning would result from increasing amounts of practice on the first task. This hypothesis should be tested under better controlled conditions than obtained in the present study. Unfortunately, association reversal has not been used as a condition under which variables such as the amount of first task practice have been systematically studied. Data bearing upon this hypothesis are, therefore, not available. This interpretation does, however, receive support from the studies of transfer under other conditions (8).

Type of work experience. Two comparisons are of interest under this rubric. One of these is that provided by the *F* ratios in Table 3 for Paradigm 3 and the other concerns the transfer effects of the two types of familiarization training.

The *F* ratios obtained from the variance estimates based upon groups with different types of experience obtained with Paradigm 3 (Table 3) were not significant. This indicates that the three different types of Task 1 training have about the same net effects when the opportunity to practice each of them is the same.

Separate analyses of the transfer effects resulting from dock experience (ED group) related to two conditions of familiarization training. The dock (ulstrith 1945) primarily provides the mechanic with an opportunity to learn the names of defects (i.e., the C terms in the S-C and C-S problems). In technical school C-S associations are the transfer task which is practiced more than S-C associations for reasons already indicated. Therefore, in the former case—C-S learning—the dock experience constitutes *stimulus training.* In the other case—S-C learning—the dock experience constitutes *response training.* When the data are examined in terms of these two conditions, some interesting trends can be noted which may deserve additional study. These results are presented in Table 7 and Fig. 2. They indicate that when trainees received stimulus training a *small* amount of *positive* transfer was obtained (C-S problem scores). There was one minor exception to this, viz., when the effects were measured by the general systems problems (Form B) only .71 per cent transfer was obtained. On the other hand, a moderate amount of *negative* transfer effect was obtained with the S-C problems (general systems, Form A). This amounted to −15.16 per cent.

These results (see Table 7 and Fig. 2) suggest that when terms are unfamiliar and complex, learning them before one learns to associate them with other terms has on the one hand a facilitating effect and on the other hand an interfering effect. If they are to serve as *cue terms* in the second task, learning the terms facilitates learning the new associative connections comprising the second task; however, if they serve as *response* terms, a small amount of practice with them interferes with learning the new associative connections comprising the second task.

The facilitating effect obtained when the words served as cue terms could be interpreted as an indication that the terms initially were difficult to differentiate and that "familiarization" training served as predifferentiation training.

The negative transfer effect obtained when the words served as response terms might be accounted for in terms of either or both of two factors. One is the complexity of the associative linkages. The other may be designated selective practice. This refers to the probability that while working on the dock ("familiarization training") only *some* defects were experienced, rather than all of these which could produce a given set of symp-

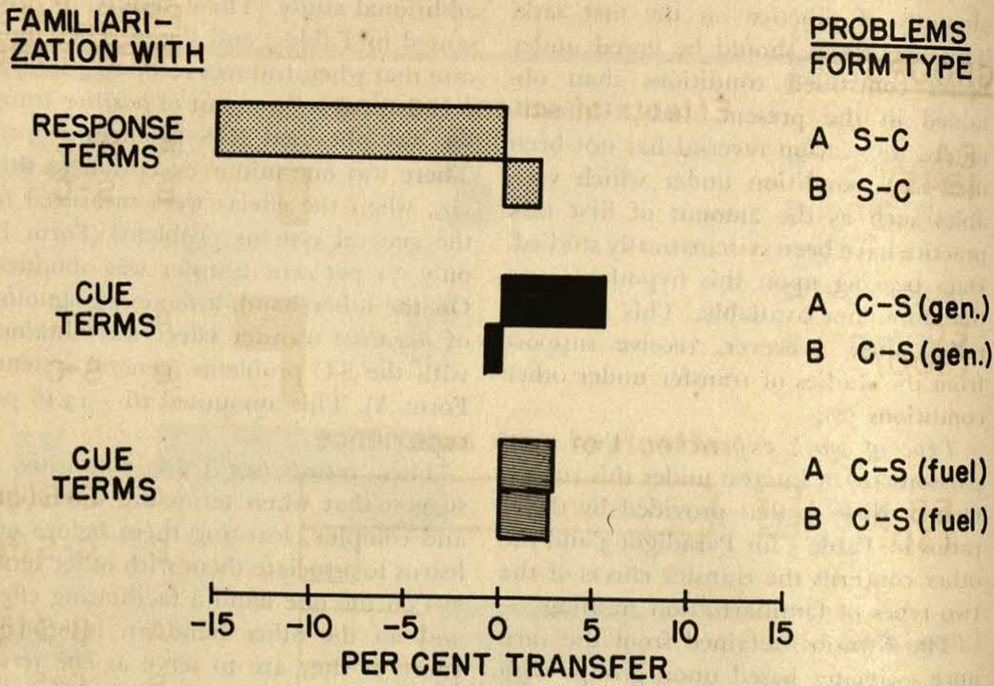


FIG. 2. Percentage of transfer effects resulting from familiarization training—Group ED, ulstrith 1945.

toms. The transfer task required the trainee to learn to associate several causes with a symptom and he may have experienced some but not all of these on the dock. As a result, in school some of the S-C relationships may have been learned better than others. If this were the case, lowered recall test scores might be due to response competition. When linkages are complexly interrelated, disproportionate learning of some terms may have the effect of raising the probabilities of these responses and, correspondingly, of lowering those for the others taught in the transfer task. The implications of these results are of sufficiently general significance to be investigated under more controlled conditions.

Retroaction Effects

The retroaction effects of learning associations in the "forward" and then in the "backward" direction can be determined from the data provided by three groups, EA (ulstrith 845), EC (ulstrith 900) and EB (ulstrith 30). These comparisons appear in the first two paradigms concerned with school-job sequence and amount of experience, respectively.

School-job sequence. Appreciable retroaction effects were produced by the two sequences studied: (*a*) school then job, and (*b*) job then school. The EB group (ulstrith 30) on the C-S problems (Form A), for example, showed the largest amount of retroactive interference, namely, -43.46 per cent. It is interesting that retroactive interference of this relative amount would be produced by nine months of job experience. One possible reason for a loss of this magnitude in the ability of mechanics to recall the C-S associations may be that they may not have learned the operation of the engine sufficiently well. Support for this interpretation comes from the relatively large number of erroneous responses made to the problems by all groups.

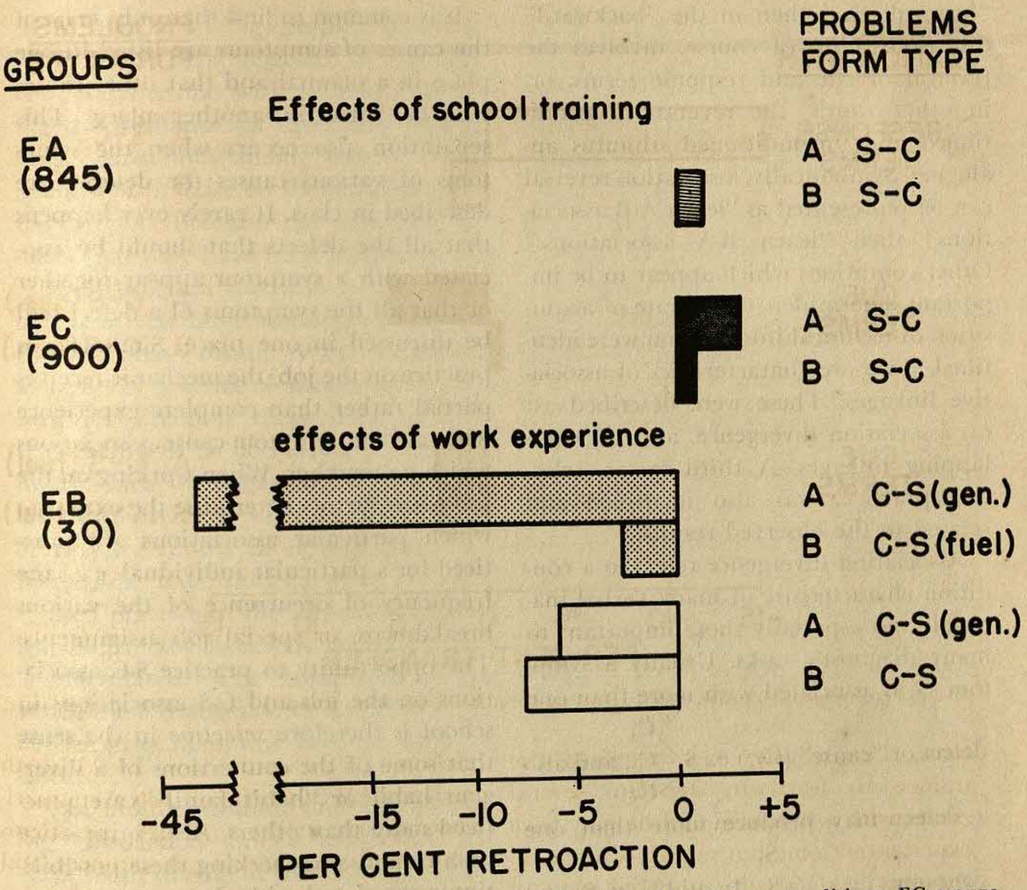


FIG. 3. Retroaction effects of association reversal as measured under two conditions. EC groups had school training after job experience whereas the EB group had job experience after completing technical school. See Table 3.

Amount of experience. The finding that with increased practice on the first task trainees show positive transfer effects under the conditions of association reversal, instead of negative effects which are obtained with smaller amounts of practice, is corroborated by the retroaction data. There were small but positive retroaction effects, i.e., *retroactive facilitation,* when the more experienced mechanics were tested for recall of S-C associations (see Fig. 3). The interpolated training received in technical school appears to have made these airmen somewhat more able than the pipeline students to recall S-C associations.

DISCUSSION

The importance of the present study appears to lie in the identification of some psychological problems that have gone unrecognized but which nevertheless have general implications for both theoretical and applied psychology. These are problems of habit interaction which appear under conditions that may be considered special cases of two familiar psychological paradigms—transfer and retroaction.

A variation of each paradigm is generated by what has been designated *association reversal,* a condition characterized by the learning of linkages in the "forward" and then in the "backward" direction. This, of course, involves the reversal of cue and response terms, or, in other words, the reversal of conditioned and unconditioned stimulus analogues. Symbolically, association reversal can be represented as "learn A-B associations" then "learn B-A associations." Other conditions which appear to be important determiners of the rate of acquisition of technical information were identified. Two are characteristics of associative linkages. These were described as: (*a*) association divergence, and (*b*) overlapping linkages. A third factor, selective practice, was also identified and related to the observed results.

Association divergence refers to a condition characteristic of many verbal materials and especially those important to many diagnostic tasks. Usually a symptom (S) is associated with more than one defect or "cause" (C), i.e., $S_1\!\begin{array}{l}\diagup C_1\\ - C_2\\ \diagdown C_N\end{array}$, and also a defect may produce more than one symptom, i.e., $C_1\!\begin{array}{l}\diagup S_1\\ - S_2\\ \diagdown S_N\end{array}$. In addition, S_1 and S_2 may both be "caused" by C_2 as well as by other causes. These relationships can be represented as follows.

$$\begin{tikzcd} S_1 \arrow[r, dash] \arrow[dr, dash] & C_1 \\ S_2 \arrow[r, dash] \arrow[ur, dash] & C_2 \end{tikzcd}$$

The lines joining S's and C's denote linkages to be formed and crossed lines represent "overlapping." Linkages between a symptom and two or more causes represents a divergent mechanism involving response sharing. Assuming that *S* learns in the S-C direction then the reverse arrangement—linkage between more than one symptom and a cause—represents a different type of divergent mechanism, namely one involving symptom sharing.

It is common to find that only *some* of the causes of symptoms are listed in one place in a manual and that other causes will be listed in another place. This separation also occurs when the symptoms of various causes (or defects) are described in class. It rarely ever happens that all the defects that should be associated with a symptom appear together or that all the symptoms of a defect will be discussed in one place. Similarly, in practice on the job, the mechanic receives partial rather than complete experience with all the symptom-cause associations which go together. When working on the job many factors determine the extent to which particular associations are practiced for a particular individual, e.g., the frequency of occurrence of the various breakdowns, or special job assignments. The opportunity to practice S-C associations on the job and C-S associations in school is therefore *selective* in the sense that some of the connections of a divergent habit or "habit family" are practiced more than others. Also, in practice each response of checking these possibilities is associated with the outcome "correct" according to a different probability schedule.

The results of the present study provide a basis for the formulation of some tentative hypotheses that could serve as a point of departure for research on the effects of various factors under the condition of association reversal. The results of this research could apply to training for a wide variety of tasks, e.g. language translation, telegraphy (code receiving and sending), trouble shooting, and medical diagnosis. Only one study of association reversal has been reported (13); however, Postman and Postman used simple paired associates rather than overlapping divergent associations, making their results of limited applicability to learning involving complex materials.

One hypothesis suggested by the present results is that with materials requiring the formation of divergent overlapping connections among terms a small amount of practice on the associations in the forward direction not only does not mean that they have been learned in the reverse direction but also it may mean that relatively large amounts of negative transfer may result. With increasing amounts of practice of associations in the forward direction these data indicate that there will be decreasing amounts of interference with learning them in the backward direction when a small amount of practice results in interference. Further, these data indicate that with increased practice in the forward direction learning in the backward direction may be facilitated. The conditions under which these changes occur and the factors determining their rate need to be identified.

The most significant finding concerning retroaction under association reversal with divergent overlapping associations was the relatively large amount of interference which was obtained. When the interpolated task was line experience, there was as much as 43.46 per cent interference with the recall of C-S associations.

Another problem to which the present data apply is the transfer effect of two different conditions of selective familiarization training to performance on tasks which require trainees to learn divergent connections. One condition is represented by selective familiarization training with cue stimuli—i.e., the C-terms in the C-S associations—and this was designated *stimulus training;* the other by selective familiarization training with response terms—i.e., some C-terms in the S-C associations—*response training.* Unfortunately when trainees were later required to use these terms, the amount of practice was not equal for the two directions of association. This condition, among others, militates against an adequate interpretation of the relative effects of these two conditions of selective familiarization training; however, moderate negative transfer effects were obtained with response training; and small positive transfer effects were obtained with stimulus training. One of the many basic questions that remain unanswered concerns the relative amount and direction of transfer effect that would be produced if the amount of practice on the transfer task were kept constant for both conditions of familiarization training.

While there were some discrepancies, the present results are sufficiently consistent to be interpreted as an indication that additional research on these problems would be useful. Certainly more research should be done before the implications which these data suggest are related to current training practices. However, if the implications of the present study should be verified, then some rather specific recommendations could be made regarding the curricula of courses designed to teach complexly interrelated verbal associations such as diagnostic information.

SUMMARY

A relatively large sample of Air Force trainees and mechanics was tested with miniature diagnostic problems and then subgroups were identified which differed in ways which were assumed to be psychologically significant. An analytical method was used which employed both a task analysis of jobs and training situations as well as Toops's addend coding procedure with respect to the biographical information provided by the examined airmen. This method, combined with the use of two types of problems, made possible: (*a*) the separation of transfer and retroaction effects of different sequences of technical school training and work experience, and (*b*) the determination of these effects under differing conditions.

One set of problems provided the airman with symptoms of malfunction and asked for recall of all causes that could produce it; the other set asked for the symptoms that could result from a particular cause of malfunction.

Each of the three paradigms employed three groups of subjects, one of which was a group of pipeline students who served as controls. The first paradigm was used to determine the effects of school-job sequences, i.e., line experience before vs. after technical school. The second determined the effects of different amounts of work experience preceding technical training, viz., less than 9 months vs. 9 to 18 months on the line; the third paradigm was used to determine the effects of different types of work experience, i.e., line vs. dock, when the amount is constant and brief.

Five of the six analyses of variance resulted in significant F ratios ($p < .01$ level) in the first paradigm; one of the six F ratios was significant in the second, and none of the F ratios was significant in the third paradigm. The school-job sequence was clearly an important factor determining the score variance for the diagnostic problems.

When the data were examined in terms of the conditions of transfer and retroaction that obtained for association reversal and familiarization training, some interesting findings were uncovered. These results, however, may apply only to learning tasks that can be described as consisting of divergent overlapping associations and under conditions of selective practice, since these were the conditions that obtained for the present study.

When trainees learned association in one direction and then in the reverse direction—*association reversal*—with one minor exception, either small or moderate amounts of *negative transfer* were found. When the mechanic was first trained to only a limited extent (less than nine months) on S-C associations while working on the line and then studied C-S associations in school, values ranged from .08 to 12.58 per cent. When they went to school first and then worked on the line, values ranged from 4.49 to 13.65 per cent. Similarly, under association reversal appreciable amounts of *retroactive interference* were found when the mechanics first learned C-S associations in school and then S-C associations while working on the job. These ranged from 2.65 to 43.46 per cent. On the other hand, *positive* transfer effects were found under association reversal when the first task was practiced to a relatively high level (i.e., from 9 to 18 months of line experience) before training in the reverse direction was undertaken. These values varied from 1.91 to 9.69 per cent. Either no interference or a small amount of improved recall resulted from a relatively large amount of interpolated practice of associations in the reverse direction of that originally learned.

With respect to *selective* "familiarization" training the results of the present study were that (*a*) stimulus training produced, with one minor exception, positive transfer effects varying from 2.56 to 5.46 per cent, and (*b*) response training produced transfer effects which were predominantly negative.

REFERENCES

1. BIRGE, JANE S. The role of verbal responses in transfer. Unpublished doctor's dissertation, Yale Univer., 1951.
2. BUNCH, M. E. The amount of transfer in rational learning as a function of time. *J. comp. Psychol.*, 1936, **22**, 325-337.
3. BUNCH, M. E. Transfer of training in the mastery of an antagonistic habit after varying intervals of time. *J. comp. Psychol.*, 1939, **28**, 189-200.
4. EDWARDS, A. L. *Experimental design in psychological research*. New York: Rinehart, 1950.
5. GAGNÉ, R. M., BAKER, KATHERINE E., & FOSTER, HARRIET. On the relation between similarity and transfer of training in the learning of discriminative motor tasks. *Psychol. Rev.*, 1950, **57**, 67-79.
6. GOSS, A. E. Transfer as a function of type and amount of preliminary experience with task stimuli. *J. exp. Psychol.*, 1953, **46**, 419-428.
7. LEWIN, K. *Dynamic theory of personality*. New York: McGraw-Hill, 1935.
8. MCGEOCH, J. A., & IRION, A. L. *The psychology of human learning*. New York: Longmans, Green, 1952.
9. MILLER, N. E. Theory and experiment relating psychoanalytic displacement to stimulus-response generalization. *J. abnorm. soc. Psychol.*, 1948, **43**, 155-178.
10. MILLER, N. E., & DOLLARD, J. *Social learning and imitation*. New Haven: Yale Univer. Press, 1941.
11. MOOD, A. M. *Introduction to the theory of statistics*. New York: McGraw-Hill, 1950.
12. OSGOOD, C. E. *Method and theory in experimental psychology*. New York: Oxford, 1953.
13. POSTMAN, L., & POSTMAN, DOROTHY L. Change in set as a determinant of retroactive inhibition. *Amer. J. Psychol.*, 1948, **61**, 236-242.
14. ROSSMAN, IRMA L., & GOSS, A. E. The acquired distinctiveness of cues: The role of discriminative verbal responses in facilitating the acquisition of discriminative motor responses. *J. exp. Psychol.*, 1951, **42**, 173-182.
15. SIIPOLA, E. M., & ISRAEL, H. E. Habit-interference as dependent upon stage of training. *Amer. J. Psychol.*, 1933, **45**, 205-227.
16. STOLUROW, L. M., BERGUM, B., HODGSON, T., & SILVA, J. The efficient course of action in trouble shooting as a joint function of probability and cost. *Educ. psychol. Measmt*, **15**, 462-477.
17. TOOPS, H. A. The use of addends in experimental control, social census, and managerial research. *Psychol. Bull.*, 1948, **45**, 41-74.

(Accepted for publication November 6, 1955)

Vol. 70, No. 13 Whole No. 420, 1956

Psychological Monographs: General and Applied

A Content-Analysis Method for Studying Psychotherapy[1]

EDWARD J. MURRAY
Yale University

I. INTRODUCTION

PSYCHOTHERAPY is of considerable interest to psychologists today not only because of the importance of psychotherapy in the treatment of neuroses and psychoses but also because psychotherapy is a unique source of data about some of the most important and elusive processes of human behavior. Much is yet to be learned, however, about psychotherapy itself. In spite of the general descriptions that are available, it is not clear just what goes on between a patient and his therapist. To some extent this is due to the complexity and the number of events in psychotherapy. In addition, the subjectivity of the therapist's report makes evaluation difficult. It is felt that an objective method of studying psychotherapy would increase our understanding of psychotherapy and human behavior in general.

The phenomena of greatest interest in psychotherapy are emotional in nature. The problem of how to approach these phenomena objectively is a difficult one. The observable events in psychotherapy may be grouped as physiological, gross behavioral, and verbal (10,13,23). However, it is not clear just how these observables are related to the underlying emotions and to one another. A full understanding of psychotherapy would require information on all the observables. The present study is aimed at an understanding of the verbal phenomena in psychotherapy with respect to the underlying emotional processes.

The *meaning* functions of verbal behavior were selected for study. Although there have been interesting studies on the grammatical and formal properties of the verbal behavior of a patient in psychotherapy (10), the content of the patient's speech seems more related to the major theories of personality. The general method for studying the meaning

---

[1] This study is based on a dissertation presented to the faculty of the Graduate School of Yale University in partial fulfillment of the requirements for a Ph.D. in psychology. The author wishes to express his appreciation to Professors Neal E. Miller and John Dollard for their encouragement and support. The author also wishes to thank the other members of his committee, Professors Irving L. Janis and Frank Auld, Jr. Professor Auld very generously devoted a great deal of time to the reliability studies presented here. This study was supported by a U. S. Public Health Service Predoctoral Research Fellowship. Professors Miller and Dollard also made some funds available from their respective U. S. Public Health Service grants. Professor Mark May, on behalf of the Institute of Human Relations, similarly made some funds available. The author also wishes to thank Professors John Dollard, George F. Mahl, Frederick Redlich, Merton Gill, and Richard Newman of the Departments of Psychology and Psychiatry at Yale University for providing typescripts of their psychotherapy cases and making available their fine sound equipment. A part of the cost of publication of this study was covered by a grant by Dr. Steuart H. Britt on behalf of the Britt Foundation, for which the author is grateful. Finally, the author is appreciative to his wife for her laborious work on this research as well as her faith and support.

Bureau Edn'l. Research
DAVID
Dated

properties of language is called *content analysis*. The previous use of content analysis in psychotherapy has been summarized by Auld and Murray (1). The present paper differs from previous work in two ways: first, the method was developed in a way which was thought to be most relevant to an eventual understanding of the underlying emotional processes; and second, the method was related to the systematic accounts of human behavior of psychoanalytic theory (5,7) and learning theory (3,11).

The purpose of the present study is to describe the development, reliability, and validity of an objective method of studying psychotherapy. The method is a way of obtaining measures of the manifest content of the verbal behavior of the patient, and also of the therapist, with respect to underlying emotional processes in terms of psychoanalytic theory and learning theory. The method is limited to the conscious part of therapy. In all but a few specified instances the patient's words were taken at face value. Various studies designed to establish the reliability of the method will be presented. The validity of the method is somewhat difficult to establish since there is no generally accepted criterion. However, the general basis of validity of scientific measures of this kind is their usefulness in establishing a set of lawful relationships (12). It is beyond the scope of the present paper to establish the complete range and limitations of the measures which have been developed. Nevertheless, a number of studies designed to demonstrate the usefulness of the method will be presented. These will include an example of the use of the method in the study of a single psychotherapy case which has importance in a major theoretical controversy. Furthermore, a study demonstrating the use of the technique in determining a general relationship in a group of psychotherapy cases will be described. Finally, an attempt will be made to evaluate the usefulness and limitations of the method on the basis of these results and several previously published studies. In addition, comparisons of the results of these studies will be made with what is already clinically known in a general, qualitative way.

II. The Method

A. The Development of the Method

The preliminary studies done during the development of the method will be briefly reviewed and then the method as it was finally used in the present research will be described in detail.

The first study involving the method has been reported in detail in an earlier paper (18). A few simple categories were devised to describe the events in seventeen consecutive, tape-recorded psychotherapy hours with one patient. The patient categories were hostility and defenses. Hostility was loosely defined as the expression of any feeling of frustration, resentment, or anger. The hostility category was divided into six subcategories of hostility according to the referent: *(a)* mother, *(b)* aunt, *(c)* other people, *(d)* general situations and groups, *(e)* the therapist, and *(f)* a vague "at home" (referring to one or more persons in the patient's home). The defense category was composed of: *(a)* intellectual defensive statements which included the patient's views on philosophy, science, current events, etc., and *(b)* complaints about a wide variety of physical symptoms and discomforts. Everything else the patient said was considered irrelevant. The unit scored was called a statement. This was either a simple sentence or the meaning phrase of a more complicated sentence. The tape of each hour was played very slowly, and each statement was judged either as belonging to one of the eight categories or as irrelevant. A reliability study with three other judges scoring a single randomly selected hour showed that the categories were defined in a way which was explicit enough for teaching other people the method with a fair degree of agreement with the main scorer. The reliability for the eight categories was high for all three scorers.

The results of this preliminary study showed that the number of hostility statements increased during the course of the case and defensive statements decreased. This was interpreted in the following way: if it is assumed that in this case hostility was inhibited by anxiety and the defensive statements were motivated by anxiety, then it would seem that during therapy the patient's anxiety decreased and permitted the expression of hostility. The results also showed that the intellectual and physical complaint defenses operated as alternative members of a defensive hierarchy. There was also some suggestion that when anxiety is increased in the later stages of therapy, previously uninterpreted (or unpunished) defenses have a greater probability of occurrence than do interpreted (or punished) defenses.

The content-analysis results of the preliminary study also showed a phenomenon which seemed to be related to displacement (15, 16, 17, 18). The patient first expressed hostility about his mother, then about his aunt, and, finally, about other people. This is contrary to the expectation that a patient will first express hostility about the less significant and less anxiety-arousing figures in his life and then proceed to the more important people such as his mother. But on the basis of the content-analysis data, the hypothesis was formulated that a person would first approach the primary and most significant goal in a given situation and would displace to less significant but similar goals only as anxiety about approaching the primary goal increased. This hypothesis was tested and verified in an animal experiment done in collaboration with M. M. Berkun and reported on in detail in a previous paper (21). Briefly, in this experiment hungry rats were trained to run down a white alley to get to a food box. A conflict was established by shocking the rats at the food box. After this the rats would only go part way down the alley. However, at this point they climbed out of the white alley into an adjacent gray alley which they had previously ignored. In the gray alley they went farther down toward a similar goal and then climbed into an adjacent black alley. In the black alley most of the rats went all of the way down to the food box. This experiment has meen mentioned in order to show that data on psychotherapy obtained with the content-analysis method can influence theoretical thought and lead to controlled experimentation.

The results of the content analysis were meaningful enough and stimulating enough to encourage further development of the method. The task of developing a more detailed set of categories was begun on a series of tape recordings of psychotherapy hours. Immediately a difficulty became apparent: the judges were not rating the same number of units. This made comparisons of content scoring difficult and impeded the development of the categories. Therefore, it was decided to separate the tasks of unit scoring and content scoring. In order to do this typescripts of the hours were used. A special study was made to define the unit and train one person to do all the unit scoring in advance. (This is described in the next section.) The content scoring was then done by listening to the tape recording and using the typescripts to locate the unit.

The advantage of a *set* of categories over a single measure was emphasized at this point in the development of the method by a study done in collaboration with F. Auld and Alice White which has been published (20). In this study, a comparison was made between the Discomfort Relief Quotient and the present content-analysis method. The DRQ consists of the number of discomfort units over the number of discomfort plus relief units (2). Thus, the DRQ gives a measure of verbal tension and may be considered the simplest form of content analysis. In the case studied in the paper, the DRQ showed no change throughout therapy and seemed to indicate no progress. On the other hand, the content analysis showed a dramatic shift from a defensive cover story about the patient's mother stealing the patient's children to a discussion of conflicts with the patient's husband. Furthermore, the discussion of conflicts with the husband showed a shift from a minor hostility conflict about going to parties, vacation plans, etc., to a major sexual conflict. All of these significant changes were missed by the single dimension of the DRQ. The results of the content analysis were in line with a qualitative, clinical analysis of the same case by Dollard, Auld, and White (4).

Therefore, in order to develop a method which would adequately describe what was going on in therapy it seemed necessary to devise and define reliably a detailed set of categories. The categories eventually used were heavily influenced by theories of psychoanalysis and learning. In addition, a considerable amount of time was spent reading psychoanalytic descriptions of case material and the available verbatim scripts of psychotherapy interviews. This period in the development of the method was characterized by a great deal of trial and errror. Finally, a workable set of categories for both the therapist and patient were derived. The task of showing their reliability and validity is the purpose of the research reported in the following pages. There is no implication that the task has been fully completed nor that the development of the method has ceased.

B. The Unit of Patient Content

The purpose of this part of the study was to establish a set of rules for determining patient response units which would be teachable and highly reliable. A by-product of the study was that one highly reliable unit determiner, other than the writer, was available to unit-score typewritten hours in advance of content analysis. This eliminated any bias on the part of the writer and other content judges to find more units of one content category than another, or to find more units of a particular category at one stage of therapy than another.

The unit used was the *statement* or *meaning phrase*. The rules defining this unit appear in Appendix B. A unit similar to this has been used in work on the Discomfort Relief Quotient. Dollard and Mowrer (2) compared the reliability of the DRQ when three different units were used. A "thought clause" unit gave an intercorrelation of +.88 with 10 judges. However, the question arises, how reliable is the determination of the "thought clause" unit? Dollard and Mowrer found that the intercorrelation of 18 scorers for the number of units per page was only +.64. The absolute values ranged from 50 to 80 units on an average page. Although this had little effect on the DRQ, it would in all probability distort a more complicated set of categories. Seeman (24) used a unit which seems to be similar to the one proposed here. He found an average agreement among four judges of 95% when such units of content were counted, and 81% when such units of attitudes were counted. It can be concluded from these studies that a unit like the statement is feasible.

The procedure in this study was as follows: Six typewritten hours from a patient seen 21 times were available. These had been randomly selected for typing. The six were then arranged in a random order. A set of rules for unit scoring was devised and applied to the first two hours by the writer. The mechanics consisted in simply making a stroke after each string of words judged to be a unit. The rules were then taught to a social worker who was naive as far as this study was concerned. Hour 1 was used for teaching examples. The judge then scored Hour 2 independently and the results were compared with the writer's scoring. Discrepancies were discussed and the rules clarified. Then the judge and the writer scored Hour 3 independently and compared results. This was repeated for Hours 4, 5, and 6. In this way the rules were made more explicit and general, while personal tendencies to make errors, overlook certain things, and introduce unverbalized rules were reduced.

The results are presented in Table 1. It can be seen that in the vast majority of cases both the judge and the writer

TABLE 1

THE SCORING OF PATIENT-CONTENT UNITS IN THERAPY HOURS BY TWO JUDGES

No. of units a given set of words was divided into by—		Hours				
Judge A	Judge B	2	3	4	5	6
1	5	—	—	—	1*	—
1	4	—	2	1	—	—
1	3	—	3	3	2	2
1	2	14	42	19	16	12
1	1	412	523	532	527	446
2	1	40	7	6	23	6
3	1	2	—	—	—	2
4	1	—	—	—	—	1
Other discrepancies		3	2	12	7	6
Total instances		471	579	573	576	475
No. of exact agreements ÷ Total no. of instances		.88	.91	.93	.92	.94
Total units, Judge A		515	586	579	599	488
Total units, Judge B		485	633	601	600	491

* To be read: There was 1 instance, in material obtained in the 5th hour of tape-recording, where Judge A counted a given set of words as 1 unit and Judge B counted the same set of words as 5 units.

scored precisely the same string of words as a unit. This agreement was high in the beginning but improved with practice and discussion. The bulk of the discrepancies consisted in one scorer's dividing a given string of words into two or three units while the other considered it just one unit. Thus there were few discrepancies where there was no correspondence between the units. An analysis of the discrepancies shows no tendency for making discrepancies with one kind of content such as hostility or physical complaints. Rather, most of the discrepancies were due to one grammatical type: the relational sentence. The "other discrepancies" consisted of (*a*) a word or phrase put with the preceding sentence by one and with the following sentence by the other; (*b*) phrases considered unscorable by one but scored by the other; (*c*) a string of words divided into three units by one and two by the other; (*d*) minor discrepancies. Discrepancies of all types decreased considerably by the sixth hour. The last line on Table 1 shows the total number of units scored by the judge and the writer. It can be seen that the totals became closer as the training went on.

The reliability of this method seems intuitively high. Yet it is a little difficult to express the agreement statistically since there is no total number of units known beforehand. However, one way of treating the results is to assume that strokes could be put after every word. One may then determine the number of words after which both scorers put strokes, the number after which only A put a stroke, the number after which only B put a stroke, and the number after which neither put a stroke. This was done for Hour 6. The resulting 2×2 table yields a phi coefficient of $+.96$. The chi square associated with this is 4155, which is decidedly beyond the .001 level.

C. *The Patient-Statement Categories*

The system of patient-content categories was based on the concepts of motivation and conflict. The method attempted to order the content of the patient's speech so that it would be relevant to an eventual understanding of the underlying drives of the patient without assuming that the relationship is necessarily a simple one. There is no doubt, for example, that a patient may express a drive for sex when it is really his dependency needs that he wants satisfied. However, before making such an interpretation it would seem desirable to have an objective record of the patient's *expressed* needs. Therefore, the system was defined in terms of manifest motivations.

Four main drives were considered in the system of categories: sex, affection, dependence, and independence. Sex referred to all erotic needs including sexual play, intercourse, masturbation, and perversions. It also referred to sexual curiosity and interest, as well as bodily functions related to sex. Affection referred to needs for love, human warmth, friendship, acceptance, etc. Dependence referred to needs to be taken care of, helped, nurtured, etc. Independence included needs to be independent, self-assertive, mature, successful, and competitive. This independence category was perhaps too broadly defined. Statements by the patient which referred to drives not included under sex, affection, dependence, or independence, and patient statements in which the drive involved was not clearly specified were grouped together in an unspecified drive category. However, in a few cases a special drive category, such as hunger need and maternal need, was set up because of the high frequency of statements in that area by a particular patient.

Statements by the patient were first put into one of the drive categories and then put into one of three subcategories under each drive. These three subcategories were defined according to whether the patient was: 1, simply expressing a need, describing past gratifications of that need, and making plans for the future gratification of that need, or 2, describing a partial or complete blocking of that need by *internal* factors such as anxiety, fear, and guilt, or 3, complaining about and getting angry about frustration of that drive by *external,* environmental factors. These were referred to as the three components of a drive: approach, anxiety, and frustration. "I enjoyed intercourse with my wife last night," is an example of sex approach. "I get so frightened when I go to bed with a girl that I'm impotent," is an example of sex anxiety. "I get no satisfaction because she's so frigid," is an example of sex frustration. Anxiety was usually thought of as being in conflict with the approach component of a drive. The frustration category included most of the hostile and complaining material expressed by the patient.

The anxiety component of each drive was originally subdivided into four parts: anxiety, negative to the self (feelings of inadequacy, etc.), physical symptoms, and guilt about hostility related to the drive. These four parts proved to be easily confused and separately unreliable. Therefore, they were combined. However, it may be that in the future these various aspects of anxiety may be reliably and profitably distinguished.

Several special categories were used in all cases. These were: disturbances of free association, simple agreement with the therapist, simple disagreement with the therapist, intellectual defenses, and generalized anxiety. Physical complaints and symptoms were included in the generalized anxiety category because they proved to be low in frequency and also difficult to distinguish from generalized anxiety.

Each statement, in addition to being placed in a content category, was rated as to the referent, e.g., self, mother, uncle, therapist, etc. The reliability for these judgments was extremely high in all cases. See Section IV for the reliability of statements related to the therapist.

The various categories may be combined for various purposes of analysis since they were defined in such a way as to be comparable. Thus, sex frustration, affection frustration, dependence frustration, independence frustration, and drive-unspecified frustration may be combined to give an over-all measure of frustration. The definitions and examples of all the categories are presented in Appendix C.

In order to test the reliability of the set of categories, the following procedure was used: A sophisticated psychologist was trained on about ten psychotherapy hours which had been previously scored by the writer. Conferences were held after each hour was scored and discrepancies were discussed. The reliability person then scored 15 psychotherapy hours independently and the hourly totals in each category were compared. The hourly totals were the appropriate measures for comparison since the results to be presented in Section III and IV are in terms of hourly totals. Therefore, the reliability is in terms of data pooled for the hour and not in terms of exact agreement between judgments of individual patient statements. The sample of 15 psychotherapy hours consisted of 8 randomly selected hours from Case F, 3 randomly selected hours from Case E, and the four even hours from Case B. These cases are described in Section IV. Only the first 300 patient units were scored in each hour in order to cut down the size of the task and permit a greater number of hours to be scored. Three hundred units comprised roughly three-fourths of each hour and seemed representative of the whole hour. These hours were randomly coded within each case so that the scorer was not influenced in his judgments by a knowledge of the position of the hour in the case. This coding procedure was followed in all of the scoring by the writer as well. The reliability person scored the hours while simultaneously listening to the tape recording and reading the pre-unit-scored verbatim typescripts.

Table 2 shows correlations based on 15 hourly totals (*df* = 13) for the basic categories. Approach, anxiety, and frustration each refer to a summation of these components for all drives. Sex, affection, etc., refer to a summation of the components of each drive. It can be seen that the correlation coefficients are moderately large and highly significant. However, the correlations account for only 66% of the variance on the average. Thus, there is still considerable room for improving the definitions of the categories and the method of teaching them. The reliability of the subcategories for each drive is shown in Table 3. Two of the subcategories, Dependence-Frustration and Independence-Frustration, showed little interjudge agreement and were excluded from general use. The other subcategories have moderately large, highly significant correlations. On the average, these correlations account for 62% of the variance. In general, the reliability of the patient-content categories seems adequate although not extremely high.

TABLE 2

THE RELIABILITY BETWEEN TWO JUDGES OF THE BASIC PATIENT-CONTENT CATEGORIES BASED ON FIFTEEN HOURLY TOTALS

Category	*r**	*p***
Drive component—all drives:		
Approach	.89	.001
Anxiety	.77	.001
Frustration	.78	.001
Drive—all components:		
Sex	.81	.001
Affection	.86	.001
Dependence	.58	.05
Independence	.69	.01
Other or unspecified	.92	.001
Special categories:		
Disturbance of free association	.67	.01
Agree therapist	.98	.001
Disagree therapist	.86	.001
Intellectual defenses	.82	.001
Generalized anxiety	.91	.001

* *r*, in this table and following tables, stands for the Pearson Correlation Coefficient.

** *p*, in this table and following tables, stands for the probability that the corresponding correlation coefficient is greater than zero. Unless otherwise stated this value is for both tails of the probability distribution.

TABLE 3

THE RELIABILITY BETWEEN TWO JUDGES OF PATIENT-CONTENT SUBCATEGORIES BASED ON FIFTEEN HOURLY TOTALS

Subcategory	*r*	*p*
Sex approach	.58	.05
Sex anxiety	.94	.001
Sex frustration	.63	.02
Affection approach	.90	.001
Affection anxiety	.69	.01
Affection frustration	.55	.05
Dependence approach	.84	.001
Dependence anxiety	.93	.001
Dependence frustration	.01	ns
Independence approach	.93	.001
Independence anxiety	.62	.02
Independence frustration	.03	ns
Other or unspecified anxiety	.88	.001
Other or unspecified frustration	.94	.001

D. The Therapist-Remarks Categories

While the main interest in developing the method was in the patient-content categories, it was thought desirable to devise at least a preliminary set of categories for therapist remarks, so that something would be known about the stimuli presented to the patient. A simple unit was selected for this purpose: everything a therapist said in between patient statements. However, there were a few cases where therapists and patients interrupted one another and in these instances the entire remark of the therapist was considered one unit.

The therapists' remarks were grouped into two general classes: active and passive remarks. The active remarks included all clearly interpretative, evaluative, and manipulative responses. The subcategories of active remarks consisted of: instructions to free associate, labels, discriminations, similarities, strong approvals, demands, and directions. These subcategories were drawn from the learning theory analysis of psychotherapy by Dollard and Miller (3) and from the examination of verbatim typescripts. The subcategories are defined along with examples in Appendix D.

Passive therapist remarks refer to those remarks which are primarily designed to acknowledge that the patient is talking. The subcategories consisted of: mild probings, mild approvals, and "Mm." These are also defined in Appendix D. All other therapist remarks were scored as irrelevant.

A preliminary study showed that with training even a relatively psychologically naive person could score therapist remarks reliably. However, the main reliability study was done on the same sample of psychotherapy hours ($df = 13$) used for the reliability of the patient-content categories. The same psychologist scored the therapist remarks as he scored the patient statements.

Active and passive therapist-remarks categories had moderately large, highly significant reliability coefficients as is shown in Table 4. These coefficients account for an average of 64% of the variance. The subcategories all had significant, although not universally large, correlations, except for the subcategory of mild probes. This subcategory, of course, was not used by itself for any analysis of the results.

Since the total number of therapist remarks in any one hour was not predetermined, a comparison of the totals shows the reliability of determining therapist-remark units. This is highly reliable, as is shown in Table 4.

TABLE 4

THE RELIABILITY BETWEEN TWO JUDGES OF THERAPIST-REMARKS CATEGORIES BASED ON FIFTEEN HOURLY TOTALS

Category	*r*	*p*
Active therapist remarks:	.71	.01
Instructions	.83	.001
Labels	.59	.02
Strong approvals	.95	.001
Disapprovals	.70	.01
Demands	.54	.05
Directions	.97	.001
Passive therapist remarks:	.88	.001
Mild probes	.46	ns
Mild approvals	.82	.001
Mm	.99	.001
Not classifiable	.64	.01
Total therapist remarks	.998	.001

The subcategories of discrimination and similarities had no frequency in the reliability sample. Therefore, no correlation could be computed. These categories also had very little frequency in all of the cases studied to date. This may be due to the fact that relatively short-term cases were used. In a longer, analytic therapy case, it would be expected that these categories would be more useful. Their reliability remains open to question until tested. The remaining categories in Table 4 seem to be sufficiently reliable to warrant using them.

III. THE USE OF THE METHOD IN A CASE STUDY

A. The Effect of Subtle Directiveness in Psychotherapy

The use of the content-analysis method in a case study will be illustrated by an analysis of the effects of subtle directiveness by a therapist on changes in the verbal behavior of his psychotherapy patient. From the point of view of learning theory (3, 11) it would be expected that mild approval and disapproval by the therapist would be an effective method of subtly directing a patient. It would be anticipated that patient categories which contained statements which were followed by therapist remarks having mild approval value would show an increase in frequency during the course of psychotherapy. On the other hand, it would be expected that patient categories which contained statements which were followed by mild disapproval by the therapist would show a decrease in frequency during the course of psychotherapy.

The case studied was that of Herbert Bryan, which was printed verbatim in a book by Carl R. Rogers (22). This case is a good one for studying the effects of subtle directiveness in psychotherapy because the patient was treated by the client-centered technique. The philosophy behind this technique is that the therapist must recognize the personal integrity and capacity for maturity of the patient. The client-centered therapist enables the patient to reach the level of adjustment that the patient himself desires by being accepting, understanding, and permissive (22, 23, 25). The very nature of the client-centered technique precludes any strong directiveness which, if present in the case, would probably obscure any possible effects of subtle directiveness. On the other hand, a certain amount of mild approval and restrained interpretation is recognized as occurring in client-centered therapy (22, 23) and accepted as a legitimate part of the client-centered technique (25). Therefore, since the case would be expected to have little or no strong directiveness and some mild approval and interpretation, it would seem to be an appropriate case for studying the effects of subtle directiveness by the therapist on changes in the verbal behavior of the patient.

B. Method

The content-analysis method described in Section II and Appendices B, C, and D was applied to the eight therapy hours printed in *Counseling and Psychotherapy* (22). The categories used in this study are briefly described as follows: The category of *sex* contained statements about sexual attraction to the mother, hostility to the mother for punishing the patient for talking about sex, etc.; sexual feelings towards various girl friends; expressions of feeling feminine in a voyeuristic situation, jealousy of boy friends who leave him for girls, and other homosexual expressions requiring a minimum of interpretation. *Intellectual defenses* consisted of philosophical discussion and extremely abstract speculation. *General anxiety* included statements of tension, conflict blocking, and physical sensations of various kinds. *Independence* involved several subcategories of going ahead in business, general assertiveness, freedom from the therapist, etc. *Independence*

TABLE 5

THE RELIABILITY BETWEEN TWO JUDGES OF THE CONTENT CATEGORIES USED IN THIS STUDY BASED ON FOUR HOURLY TOTALS

Category	r	p
Independence	.99	.02
Independence anxiety	.96	.05
Intellectual defenses	.36	ns
All sex	.998	.01
General anxiety	.97	.05

anxiety was the expression of fear and guilt about independent and assertive actions.

Table 5 shows the reliability of the categories used in this study. One clinical psychologist was trained on the four odd-numbered hours in the case and then his agreement with the writer determined on the remaining four hours. The reliability for each category was based on the total number of units judged in that category in the first three hundred patient units in each of the four hours ($df = 2$). The reliability was good for all categories except intellectual defenses. The intellectual defense category was retained on a highly tentative basis in this study in spite of the fact that it did not reach significance for two reasons: 1, the reliability scorer agreed with the writer on a general downward trend and 2, the category was highly reliable in the larger reliability study (see Section II).

The therapist-remarks categories which seemed most relevant to the problem were *approval* and *disapproval.* All other categories were grouped together as *neither.* Table 6 shows the agreement between the writer and the same clinical psychologist who judged the patient content in independently judging therapist approval and disapproval related to each of the patient categories. The sample consisted of all the therapist's remarks made during the first 300 patient units for each of the four odd-numbered hours. This was about 60 therapist remarks on each of the hours and added up to a 40% sample of all therapist remarks. It can be seen that both scorers agreed that *independence* was approved and never disapproved; that *independence anxiety, intellectual defenses,* and *sex* were disapproved and never approved. There is general agreement on the other categories as well.

TABLE 6

A COMPARISON OF JUDGMENTS OF THERAPIST APPROVAL AND DISAPPROVAL BY TWO JUDGES, AS RELATED TO ALL PATIENT-CONTENT CATEGORIES

Patient content categories	First judge		Second judge	
	Approvals	Disapprovals	Approvals	Disapprovals
Independence	41	0	27	0
Independence anxiety	0	9	0	2
Dependence	0	9	0	4
Dependence anxiety	0	0	0	0
Intellectual defenses	0	4	0	3
All sex	0	7	0	5
General anxiety	1	1	0	1
Other:				
Anxiety parents	1	0	1	0
Frustration parents	0	1	0	1
Positive to self	2	0	3	0
Affection therapist	2	1	2	1
Disturbance	0	0	0	0
Irrelevant	0	1	0	0

C. *Results*

The primary problem is whether the therapist influenced the verbal behavior of the patient during therapy. In order to determine this it is necessary first to demonstrate that the therapist approved and disapproved different categories. Table 7 shows all of the approvals and disapprovals found in the case by the writer. The same pattern as in Table 6 is immediately apparent. The therapist approved independence and primarily disapproved independence anxiety. Furthermore, he disapproved dependence and approved dependence anxiety. These are two different ways of doing the same thing. In addition, the therapist disapproved sex and intellectual defenses. General anxiety was both approved (in a reassuring way) and disapproved (at several points where it dominated the discussion). The other categories are shown in Table 7 to show that few other approvals or disapprovals appeared. An examination of these categories shows a pattern which fits in with the approval and disapproval in the major categories.

TABLE 7

ALL THERAPIST APPROVALS AND DISAPPROVALS FOR ALL RELEVANT PATIENT-CONTENT CATEGORIES

Patient Category	Therapist approval	Therapist disapproval
Independence	68	0
Independence anxiety	3	15
Dependence	1	16
Dependence anxiety	2	0
Intellectual defenses	0	11
All sex	2	16
General anxiety	5	4
Other:		
Anxiety parents	1	0
Frustration parents	0	1
Positive to self	5	0
Affection therapist	2	1
Disturbance	1	1
Irrelevant	2	0
Total	92	65

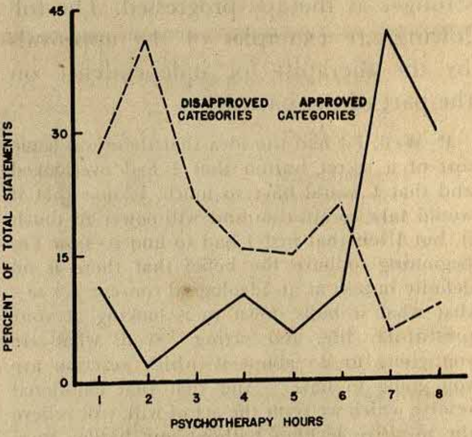


FIG. 1. The effects of approval and disapproval by the therapist on the percentage of statements in the approved and disapproved patient categories during eight hours of psychotherapy. All patient categories which were approved are pooled and all patient categories which were disapproved are pooled in this analysis.

The next step is to see whether the therapist's pattern of approval and disapproval had an effect on the patient's verbal behavior. Figure I shows the percentage of the total number of patient statements in each hour which were in the categories approved by the therapist and in the categories disapproved by the therapist. The approved categories increase as therapy progresses while the disapproved categories decrease.

The number of approvals the therapist gave on a given hour was highly correlated with the percentage of the rewarded categories ($r = .97$, $p < .01$). This was also true of disapprovals and the percentage of the disapproved categories ($r = .80$, $p < .05$). Therefore, the behavior of the therapist was consistent with the appearance of the categories.

The therapist reacted to the patient's expression of desires to be mature, independent, and self-assertive with approval on Hours, 1, 3, 4, 5, 6, 7, and 8. The approvals grew more frequent and stronger as therapy progressed. The following are examples of the approvals by the therapist for independence on the part of the patient:

P—Well, I'd had the idea that there was some sort of a secret button that I had overlooked and that I would have to touch. I knew that it would take motivation and will-power to touch it, but I felt that first I had to find it—now I'm beginning to have the belief that there is no definite button as an ideological concept *per se*—that what it boils down to is looking at your present-day life, and saying, "Well, what are you going to do about it—which reaction are you going to have?" And that that emotional resolve which we term the act of will, will relieve the negative feelings without any hidden mystery or anything like that having to be exposed.

T—That's what I think, too.

P—I'm glad to hear that.

T—Well, there's no doubt that you are right when you first came in that you had a lot to say today.

P—Oh, I knew it. I mean—I know myself pretty well; I think I'm pretty honest with myself.

T—And growing increasingly honest, I would say.

P—M-hm. Yes. I imagine that that's the true function of psychoanalysis, is to—well, to sort of force the issues, to make them clear-cut, and so that the resolve can take the right direction.

T—Well, you raised the question last time whether I thought you were making progress. I guess you wouldn't have to ask it today (Laugh).

and,

P—My work had a setback. The camera needs some more repair. It won't be ready till next week, but I think I can get ahold of another one. I notice I made an effort to locate another one, rather than wait till next week, till the first one gets fixed.

T—Do I understand that you really wanted to get to work on that?

P—Yes, I think that that would be the indication there. See, I could have waited till next Wednesday, until my own camera gets fixed, but I went this afternoon and got one lined up from one of the downtown supply houses.

T—That really means something.

P—Yes, I thought that it did. I went over and talked to the manager quite a lot. It needed a good deal in the realm of persuasion there—it was a very unusual request. I felt that I talked him into completing negotiations so that I won't have to wait till next week.

T—And that makes you feel pretty good, I presume?

P—Oh, yes. You see, I believe that some weeks ago I would have said, "Well, I believe I'll wait till Wednesday," and just seized the pretext to remain inert, but—

T—Something makes it different now?

P—Yes, there is a barrier there, but not an insurmountable one. I walked in, anyway, and started talking to the manager. (Pause.)

T—That sounds like quite a step.

As can be seen in Figure 2, the first effect of this behavior of the therapist was to raise the anxieties of the patient about *independence*. Some mild disapproval for *independence anxiety* was expressed by the therapist on Hours 5 and 6. Then, rather suddenly, as shown in the upper half of Figure 2, *independence* increased to dominate Hours 7 and 8 and was somewhat strongly approved by the therapist.

The categories of *dependence* and *dependence anxiety* had very low frequencies which did not even permit a relia-

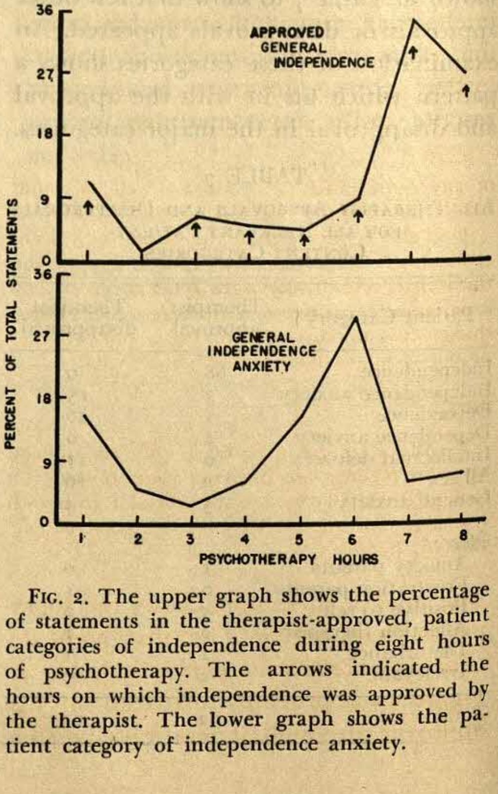


FIG. 2. The upper graph shows the percentage of statements in the therapist-approved, patient categories of independence during eight hours of psychotherapy. The arrows indicated the hours on which independence was approved by the therapist. The lower graph shows the patient category of independence anxiety.

bility coefficient to be computed. They must be regarded as unreliable and are therefore not presented. The effects of the therapist's disapproval for *dependence* and approval for *dependence anxiety* (essentially an expression of denying dependence) can be thought of as operating indirectly on the categories of *independence* and *independence anxiety.* Approving *dependence anxiety* is roughly equivalent to approving *independence.* Disapproving *dependence* is roughly equivalent to disapproving *independence anxiety.* This situation can be viewed as a double approach–avoidance conflict (14) with independence and dependence as the two goals.

At the same time that the therapist was approving *independence,* he was mildly disapproving intellectual defenses and the various sex-categories. During the first few hours of therapy the patient discussed his problem in extremely abstract intellectual terms and frequently lapsed into philosophical speculation. The following excerpts illustrate the therapist's reaction to the intellectual defense.

> P—Of course, perhaps that's too general to be of any significance. It wouldn't—well, it could hardly be anything else but a fear, could it?
>
> T—Well, I think that the more significant elements are what you experience, not what you think intellectually it might be. I mean, if you experience fears in your nightmares—that's something. Any amount of thinking about whether or not it might be a fear is . . .

and,

> T—And it is going to be a question of—you say your feelings are blocking yourself—can you look at them? Can you really consider your own reactions—not an intellectualized abstracted picture of them?

The tone of these remarks by the therapist is disapproving. The subsequent decline in *intellectual defenses* shown in Figure 3 was probably related to this disapproval although it is always possible that other factors were also operating. Incidentally, the reliability scorer who disagreed with the author on some of the minor fluctuations in the intellectual defense curve also obtained a result showing a general decline from the first half to the second half of therapy.

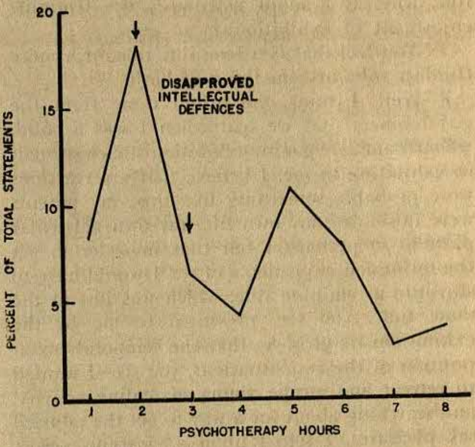


FIG. 3. The percentage of statements in the disapproved patient category of intellectual defenses during eight hours of psychotherapy. The arrows indicate hours on which intellectualization was disapproved by the therapist.

At the beginning of therapy the patient felt that his problems were primarily sexual. The therapist disagreed with the patient and at one point said that the sexual problems were only lesser manifestations of a more fundamental problem of self-respect and maturity. Many people would agree with the therapist. However, the therapist's attitude about sex seemed to influence his clarifications so as subtly to discourage the patient's dwelling on that topic. The following excerpt illustrates one of the subtle techniques used by the therapist.

> P—I vicariously experience the girl as she exhibits to me and get a vicarious sexual enjoyment from that, so that I am both the exhibitor and exhibitee. So what it seems to be is that, when my condition becomes worse, I find that my man role seems to be blocked, so I escape into what we might call a semifeminine

Bureau Ednl. Psy. Research
DAVID H TRAINING COLLEGE
Dated

role, whereas I adopt vicariously the feminine enjoyment of exhibitionism to me.

T—You feel that is a feminine role, or a more childish role, or something of both?

P—Well, I think the origin came from the girl roomers that we had when I was a child, whom I am sure got an exhibitionistic enjoyment in exhibiting to me. I believe that my reactions were probably something like this: my parents were inhibiting my own life—my own particular pattern—so perhaps I felt that in order to get the forbidden enjoyments of sex I would have to slip into a feminine role, which was just at the same time, you see, presented to me by the exhibitionistic girls. So that the temporal juxtaposition of the two situations, you see—I wanted to retreat and escape young masculine activity, and searching about for a way to get the tabooed sex pleasures, I seized upon a vicarious enjoyment of the exhibitionistic girls.

T—Let me see if I get what you're meaning there—in other words, when you're faced with something—I'm not quite sure whether the blocking is that there is something pretty difficult in this masculine role, or whether you're referring to other kinds of blocking—at any rate, when you're blocked in being a man, you tend to fall back into a possibly more feminine kind of satisfaction—certainly a more childish kind of satisfaction in voyeurism.

It seems likely that this sort of behavior on the part of the therapist contributed to the decline of the sex categories shown in Figure 4.

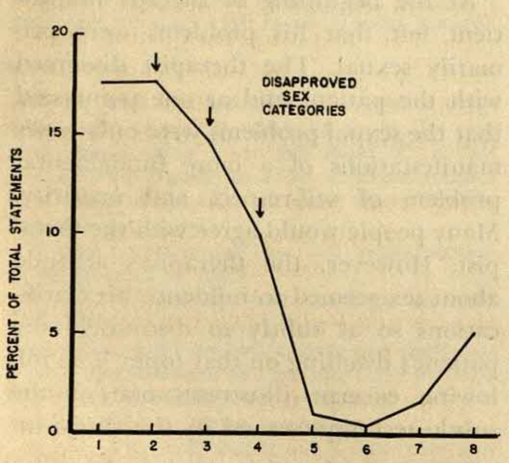


FIG. 4. The percentage of statements in the disapproved patient categories of sex during eight hours of psychotherapy. The arrows indicate hours on which discussion of sex was disapproved by the therapist.

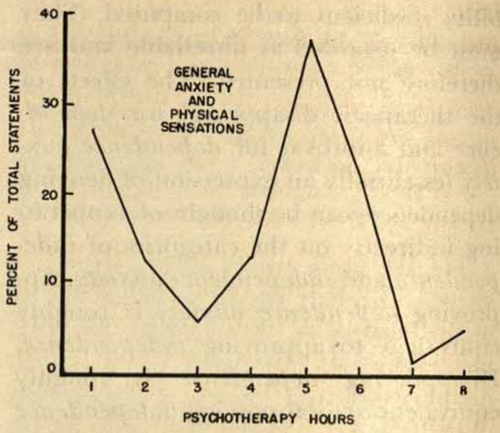


FIG. 5. The percentage of statements in the patient category of general anxiety and physical sensations during eight hours of psychotherapy.

General anxiety and physical complaints were to a large extent simply accepted by the therapist although some aspects were very mildly approved and some very mildly disapproved. This general anxiety category showed a large temporary increase on Hours 5 and 6 after an initial decline as can be seen in Figure 5. Hours 5 and 6 were probably anxiety-provoking for the patient because it was then that the major decision in the therapy was discussed. It can also be seen in Figure 3 that the intellectual defenses also increased during Hours 5 and 6. In a previous paper (18) similar data were presented and the hypothesis formulated that when anxiety is increased in the later stages of therapy all defenses may increase but previously disapproved defenses do not increase as much as defenses which have not been disapproved. This is what seems to have happened here. By Hour 5 intellectual defense and sex categories had decreased possibly as a result of the disapproval of the therapist. By Hour 7 the patient was primarily expressing the independence which had been consistently approved by the therapist. The decision to be independent was made between Hours 6 and 7. It would seem that the therapist's pattern of approval and disapproval played some role in this decision.

D. Discussion

The therapist in the case of Mr. Bryan seemed to influence the verbal behavior of the patient by subtly approving independence and self-assertion and mildly disapproving intellectual defenses and discussion of sexual problems. The therapist appeared to play an important role in the major decision of the patient. The subtle directiveness observed in this case does not consist of isolated instances of little consequence. The instances of subtle directiveness combine to form a pattern which could account for the major events in the case. The results are in agreement with the learning theory expectations and tend to support the usefulness of learning theory in understanding some of the complex events in psychotherapy.

The results of the content analysis presented in this chapter are supported by two minor studies described in Appendix A, which involved other techniques. In the first study a group of clinical psychologists showed high agreement on the approval or disapproval value of selected therapist remarks from the case of Herbert Bryan. In the second study a group of relatively naive subjects read one hour from the case of Herbert Bryan and agreed in a questionnaire that the therapist subtly got across his preference for the patient's deciding to become independent.

In the present study, the measure of directiveness was relatively crude. There are many more subtle forms of direction in this case, and in many other cases, which are somewhat difficult to measure. For example, the interest or lack of interest shown by the therapist in an item brought up by the patient may have important consequences. One way a therapist can do this is by selectively repeating one of the patient's sentences and ignoring others. This is used as a technique by some therapists for getting at important material (3). The therapist can also show interest by the selective use of "Mm" responses. Greenspoon (9) found that the selective use of "Mm" produced an increase in frequency of classes of words in a free association experiment. Finally, the therapist may express a very strong opinion by the way in which he clarifies and formulates a problem and by the language he uses. In the present case, the therapist formulated the patient's problem as a "choice–between going ahead which involves responsibility–or slipping back into the easier possibility–." Note the use of words with positive prestige value in middle class American culture such as "going ahead" and "responsibility" for one choice in conjunction with words having negative prestige value like "slipping back" and "easier possibility" for the other choice. In a middle class patient, such as Mr. Bryan, this technique might be rather powerful.

IV. THE USE OF THE METHOD IN DETERMINING A GENERAL RELATIONSHIP

A. The Patient's Feelings Toward the Therapist

The feelings that a patient has toward the therapist have been considered important in psychoanalytic theory for a long time (6). Originally the transference aspects of the patient's feelings were emphasized; certain responses by a patient to the therapist were thought of as carried over from unresolved problems with the significant people in his early life. Today analysts consider all aspects of the patient's feelings toward the therapist important, whether they are based on transference or reality (8). As the therapeutic relationship is established, the therapist can make the patient aware of his inappropriate reactions and enable him to respond realistically. In addition, the positive feelings toward the therapist are thought to keep the patient in therapy in the face of anxiety. However, there is much that is not known about the therapeutic relationship—especially its development and control. Therefore, this would seem to be an appropriate area for research.

The content-analysis method was applied to certain aspects of the therapeutic relationship in order to investigate the general problem of how the manifest attitudes of the patient toward the therapist change during therapy. The study was limited to the verbal expression of feelings toward the therapist by the patient and the therapist's verbal reactions to this expression. Data were gathered on several specific questions: Does the frequency of the patient's references to the therapist increase as therapy progresses? Is there a decrease in the verbal expression of negative feelings and an increase in the expression of positive feelings toward the therapist as therapy progresses? Or is there an increase in the expression of all kinds of feelings about the therapist? How does the verbal behavior of the therapist influence the patient's statements about the therapist?

It is quite apparent that these questions about the therapeutic relationship are limited in scope. They barely scratch the surface of the complex phenomena of the patient-therapist interaction. They are confined to several relatively simple verbal changes. However, they represent the beginning of an inquiry into that part of the area in which the content-analysis method is most applicable. Care will be taken to avoid any undue extrapolation from the verbal data obtained to the underlying emotional processes.

B. Method

A total of 110 psychotherapy hours from seven cases were used in this part of the study. Cases A, C, E, F, and G were tape recorded and transcribed by the Departments of Psychology and Psychiatry at Yale University. Cases A, E, and F contained 5, 20, and 33 consecutive hours. Case C, reported elsewhere (20), contained six randomly selected hours from a total of 17 psychotherapy hours. Case G contained every third consecutive hour over a series of 69 hours—making a total of 23 hours used. Case D, reported in detail elsewhere (18), contained 17 consecutive tape-recorded but not transcribed hours. Case B contained 8 consecutive hours printed verbatim in the book by Rogers (22).

The sample is heterogeneous with respect to the length of therapy, experience of the therapist, and theoretical orientation of the therapist. The patients were all neurotic but had various kinds of complaints and character structures. However, the cases were alike in that each consisted of an interaction of two people, one of whom came to the other for psychotherapeutic help.

The patient-content categories used in this part of the study were all those related to the therapist. The opening-and-closing-the-hour salutations and expressions of difficulty in free associating were not included. The reliabilities of the categories related to the therapist are shown in Table 8. The correlations are based on a comparison of the writer and

TABLE 8

THE RELIABILITY BETWEEN TWO JUDGES OF PATIENT-CONTENT CATEGORIES RELATED TO THE THERAPIST, BASED ON FIFTEEN HOURLY TOTALS

Category	*r*	*p*
All categories pooled	.97	.001
All positive categories	.92	.001
All negative categories	.97	.001
All anxiety categories	.92	.001
All frustration categories	.96	.001
All sex categories	—	—
All affection categories	.96	.001
All dependence categories	.88	.001
All independence categories	.54	.05

another psychologist for the first 300 patient units in each of 15 psychotherapy hours (*df* = 13). The sample included 8 randomly selected hours from Case F, 3 randomly selected hours from Case E, and the 4 even-numbered hours from Case B. The reliability scorer had been trained on the four odd hours from Case B and about six other randomly selected hours from Case F. The reliability remains high when the three cases are considered separately. The main measure used in this part of the study appears in Table 8 under the title "all categories pooled" and is highly reliable. "All positive categories" refers to approach components of sex, affection, dependence, and independence (see Section II). "All negative categories" refers to the anxiety and frustration components of all the drives (see Section II). The frequency of statements in the sex categories was too low in the reliability sample to permit the computation of a correlation. Generally, the patient-content categories related to the therapist seem to be highly reliable.

Therapist remarks studied here included any response the therapist made to a patient statement about the therapist plus any probing remark the therapist made about the patient's feelings toward the therapist. Only the major categories of therapist remarks—active and passive—were used in this part of the study (see Section II). Table 9 shows the reliability of all therapist remarks, active remarks, and passive remarks related to the patient's statement about the therapist. The same reliability scorer and the same sample used for the patient content-category reliability was used here. The reliability seems adequate.

TABLE 9

THE RELIABILITY BETWEEN TWO JUDGES OF THERAPIST-REMARKS CATEGORIES RELATED TO PATIENT STATEMENTS ABOUT THE THERAPIST, BASED ON FIFTEEN HOURLY TOTALS

Category	*r*	*p*
All therapist remarks	.90	.001
Active therapist remarks	.73	.01
Passive therapist remarks	.92	.001

Cases A, B, E, F, and G were content-analyzed by the writer with the detailed set of categories described in Section II. The tape recordings were played in a completely random and coded order while the scorer read the typescript. Case B was just read. Cases C and D were content-analyzed earlier with a simpler set of categories that did not contain the present subdivisions. Therefore, these two cases will only be used for the overall measure of patient statements related to the therapist.

C. Results

The percentage of the total number of statements made by a patient in a psychotherapy hour which was related to the therapist shows an irregular but definite increase during the course of therapy as shown in Figure 6. The data were converted into percentages in order to control for variations in the total number of statements from hour to hour. Table

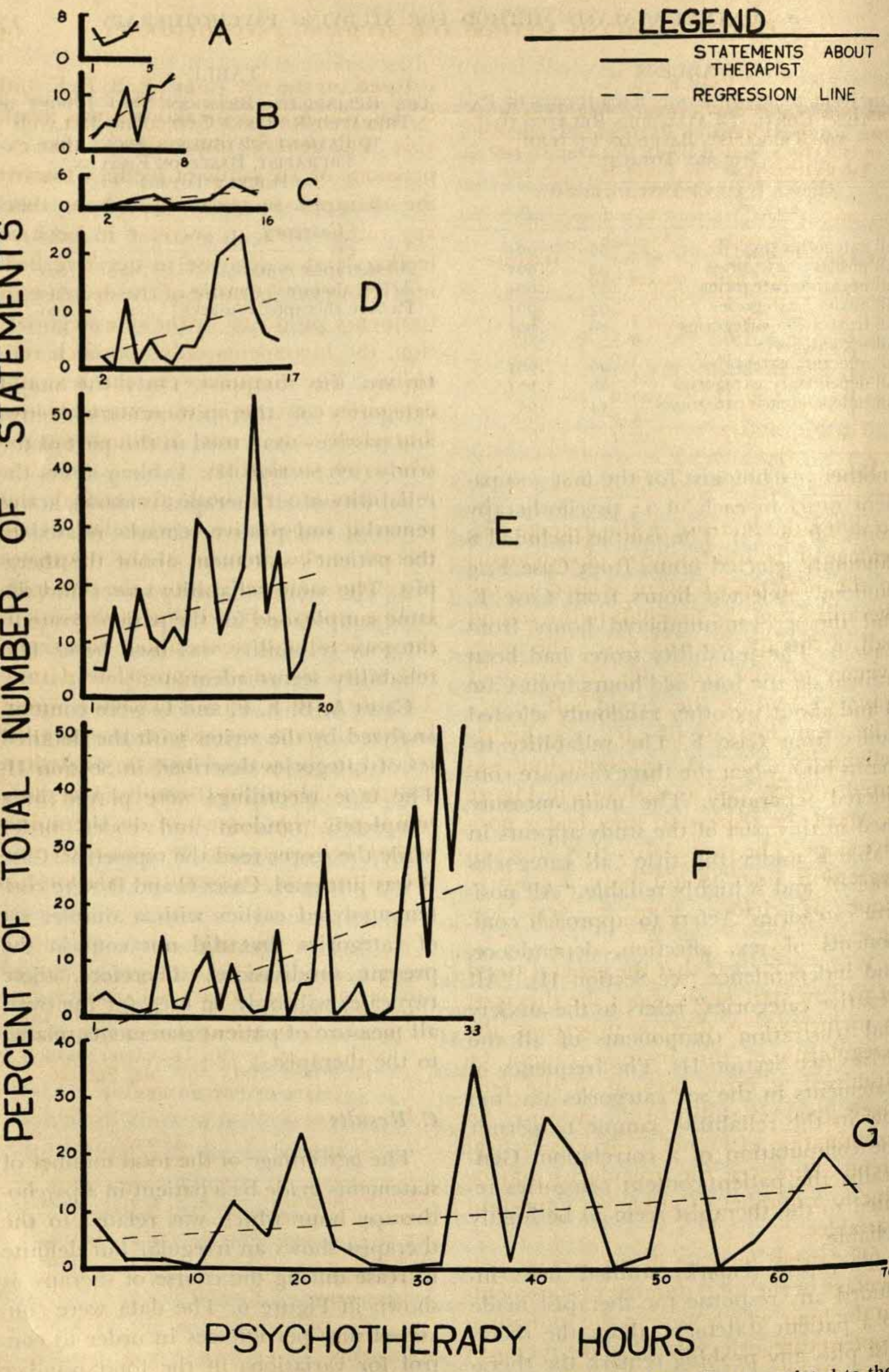


FIG. 6. The percentage of the total number of patient statements in an hour that was related to the therapist during the course of seven psychotherapy cases.

TABLE 10

THE INCREASE IN THE AVERAGE PERCENTAGE OF THE TOTAL NUMBER OF PATIENT STATEMENTS IN THE HOUR WHICH WERE RELATED TO THE THERAPIST FROM THE FIRST HALF TO THE SECOND HALF OF PSYCHOTHERAPY

Case	Average % for first half	Average % for second half	Increase
A	4.2	5.7	1.5
B	5.6	9.5	3.9
C	1.1	3.0	1.9
D	3.3	11.1	7.8
E	12.2	20.0	7.8
F	5.2	14.7	9.5
G	5.3	12.1	6.8
Mean	5.3	10.9	5.6

10 shows the increase in the average percentage of the total number of statements in the hour which were related to the therapist from the first half to the second half of therapy for each case. The mean average for the group for the first half was 5.3% and for the second half 10.9%. The mean increase of 5.6% was highly reliable ($p < .01$). A regression line was fitted to each case as is shown in Figure 6. All of the regression lines had a positive slope. The probability that all seven regression lines would have a positive slope by chance alone is very low ($p < .01$).

While it is clear that the general trend of patient statements about the therapist is to increase as therapy progresses, the irregularity of this increase cannot be dismissed as error because of the relatively good reliability of the measure. An examination of Figure 6 suggests cyclical trends. However, the irregularity may be due to a number of unknown important variables.

The mean increase in the average percentage of the total number of statements in the hour which were in the individual content categories related to the therapist from the first half to the second half of therapy is shown in Table 11. This table is based on the five cases (A, B, E, F, and G) for which such breakdown was available. The general impression is that expressions of all kinds of feelings toward the therapist increase throughout therapy rather than an increase in positive feelings and a decrease in negative feelings. However, because of the decrease in frequency produced by the subcategorization, the increments only approach reliability. The frequency in the sex and dependence categories was extremely low. Therefore this result, although suggestive, is very tentative.

A question of both theoretical and practical importance is what factors are related to the increase in the frequency of patient statements about the therapist. In this connection, the *rate* of increase in the patient statements about the therapist seems the most appropriate datum. The therapist's verbal behavior would be one of many variables expected to be related to the rate of increase of patient statements about the therapist. The problem is just what aspects of the therapist's verbal behavior can be measured and shown to have a substantial relationship

TABLE 11

THE MEAN INCREASE IN THE AVERAGE PERCENTAGE OF THE TOTAL NUMBER OF PATIENT STATEMENTS IN THE HOUR WHICH WERE IN THE INDIVIDUAL PATIENT-CONTENT CATEGORIES RELATED TO THE THERAPIST, FROM THE FIRST HALF TO THE SECOND HALF OF PSYCHOTHERAPY

Category	Average increment	*p*
Anxiety (pooled)	1.9	ns
Frustration (pooled)	2.0	.10
Sex	1.0	ns
Affection	2.0	ns
Dependence	−0.6	ns
Independence	1.1	.10
All positive	2.2	.05
All negative	3.9	.10
All statements	5.6	.01

with the rate of increase of patient statements about the therapist.

Table 12 shows, in the second column, the slope of the regression lines for the five cases on which data about the therapist's behavior were available. The next column shows the number of times the *therapist said something about the therapeutic relationship* over *the number of times the patient said something related to the therapist.* This ratio was used in order to control for the differences in the number of patient units between cases due to the variation in length of therapy. The correlation between this ratio and the slopes of the regression lines is negative but not reliable ($r = -.67$, $df = 3$).

However, the therapist's remarks were divided into two groups: active and passive. The ratio of the active therapist remarks to the number of patient units showed a positive and reliable correlation ($r = .93$, $p < .05$) with the rate of increase of patient statements about the therapist. On the other hand, the ratio of the passive therapist remarks to the number of patient units showed a reliable negative correlation ($r = -.90$, $p < .05$) with the rate of increase of the patient statements about the therapist. Furthermore, a simpler measure consisting of the percentage of therapist remarks which were active shows a highly reliable correlation ($r = .996$, $p < .01$) with the rate of increase measure. These data are all shown in Table 12. The results, then, tend to indicate that the rate of increase of patient statements about the therapist is positively related to the extent that the therapist responds to this material in an active rather than a passive way.

However, the relationship between the therapist's verbal activity and the rate of increase of the patient statements about the therapist is not restricted only to the way in which the therapist reacts to the therapist-relevant material. Rather, to some extent at least, the general verbal activity level of the therapist in dealing with all kinds of patient material—whether or not it is related to the therapist—seems to be related to the patient's verbal expression of feelings about the therapist. Thus, Table 13 shows the rate of increase of patient statements related to the therapist as a function of the way the therapist reacts to statements not related to the therapist. The measures in Table 13 are analogous to those in Table 12 but for patient statements not related to therapist and for therapist remarks about those statements. As can be seen in Table 13 the general pattern of results is similar to that in Table 12. The number of times the therapist said something

TABLE 12

THE RATE OF INCREASE OF PATIENT STATEMENTS RELATED TO THE THERAPIST AS A FUNCTION OF THE WAY THE THERAPIST REACTED TO SUCH STATEMENTS

Case	Rate of Increase (slope)	Total Ther. Rem. ÷ Pt. Units	Active Ther. Rem. ÷ Pt. Units	Passive Ther. Rem. ÷ Pt. Units	Percent Active
B	1.50	.37	.25	.12	68
F	.82	.28	.12	.16	42
E	.59	.43	.14	.29	31
A	.44	.55	.13	.42	23
G	.14	.53	.09	.44	16
Average	.70	.43	.15	.29	36

TABLE 13

THE RATE OF INCREASE OF PATIENT STATEMENTS RELATED TO THE THERAPIST AS A FUNCTION OF THE WAY THE THERAPIST REACTED TO STATEMENTS BY THE PATIENT NOT RELATED TO THE THERAPIST

Case	Rate of increase (slope)	Total ther. rem. ÷ pt. units	Active ther. rem. ÷ pt. units	Passive ther. rem. ÷ pt. units	Percent active
B	1.50	.24	.11	.14	44
F	.82	.10	.03	.07	32
E	.59	.28	.03	.25	12
A	.44	.42	.02	.40	5
F	.14	.41	.02	.39	6
Average	.70	.29	.04	.25	20

not related to the therapeutic relationship over the number of times the patient said something not related to the therapist shows a negative but not reliable correlation with the rate of increase of patient statements about the therapist ($r = -.60$, $df = 3$).

The ratio of the number of active therapist remarks not related to the therapeutic relationship over the number of patient statements not related to the therapeutic relationship is positively and significantly related ($r = .92$, $p < .05$) to the rate of increase of patient statements about the therapist. On the other hand, the ratio of passive therapist remarks to patient statements shows a negative but not significant correlation ($r = -.36$, not significant) with the rate of increase measure. Finally, the percentage of therapist remarks which were not about the therapeutic relationship which were active rather than passive showed a positive and significant correlation ($r = .94$, $p < .05$) with the rate of increase of patient statements about the therapist.

It is apparent from the above results that the rate of increase of patient statements about the therapist is a function of the degree to which the therapist is active rather than passive in responding to the patient statements in general and regardless of whether they are manifestly related to the therapist. However, the question naturally arises as to whether there are any differences in the way the therapist reacts to therapist-relevant and therapist-nonrelevant statements by the patient. The group average for each of the measures of the therapist behavior, in relation to patient statements related and patient statements not related to the therapist, are shown in Table 14. It can be seen that the therapists in the present sample responded significantly ($p < .01$) more to therapist-relevant than to therapist-nonrelevant material. This difference is due to a greater increase in active therapist remarks than to passive remarks. The active remarks showed a significant ($p < .01$) increase with therapist-relevant material while passive remarks do not. Furthermore, the percentage of therapist remarks which were active was significantly ($p < .01$) higher with therapist-relevant patient statements than with therapist-nonrelevant patient statements.

A further question which arises is whether some difference between the way the therapists reacted to therapist-relevant and therapist-nonrelevant patient statements could account for rate of increase of patient statements about the therapist. For this purpose, the differences between the percentage of active therapist remarks for the therapist-relev-

TABLE 14

The Difference Between the Way the Therapist Reacted to Patient Statements Related to the Therapist and Patient Statements Not Related to the Therapist

Measure	Average for pat. statements not related to therapist	Average for pat. statements related to therapist	Increment	p
Tot. ther. rem. ÷ Pt. units	.29	.43	.14	.01
Active ther. rem. ÷ Pt. units	.04	.15	.11	.01
Passive ther. rem. ÷ Pt. units	.25	.29	.04	ns
Per cent active	20	36	16	.01

ant and therapist-nonrelevant patient statements was correlated with the rate of increase of patient statements about the therapist. The correlation was positive but not significant ($r = .65$, not significant). More cases are needed before a conclusion is drawn about this relationship.

D. Discussion

The results tend to show that the verbal expression of the patient's feelings increases during the course of the comparatively short psychotherapy cases sampled. However, the results may be interpreted in many ways. They may mean that the patient is developing feelings about the therapist over a period of time in essentially the same way that people develop feelings about new acquaintances. On the other hand, it may be that a large part of the feelings are brought into the situation by the patient but their direct expression is inhibited by anxiety. At the very beginning of therapy they may be expressed only as displacements or may be entirely inhibited. Thus, for example, in Case D, the patient spent the first few hours in abstract intellectual discussion. In the content analysis this was scored as *intellectual defense* and properly so. But there was an undercurrent in this intellectual discussion. The patient seemed to be telling the therapist, "Look, I'm clever and I've read a lot of books. I'm worthy of your interest and respect." Later, the patient spent a good deal of time with physical complaints. Here he seemed to be saying, "I'm frightened and helpless. I need your help." Naturally these interpretations of latent meaning are very tentative but they illustrate the possibility of displaced therapist-relevant material occurring at points where the curves in Figure 6 show no direct expression of feelings about the therapist.

If the patient is expressing feelings about the therapist in a displaced way when he is talking about things not manifestly related to the therapist, it is then somewhat more understandable that the degree to which the therapist is active in his remarks about therapist-nonrelevant patient statements has an effect on the rate of increase of patient statements about the therapist.

The results then most probably mean that as therapy progresses the patient is more able to express feelings which to some extent he had all along. The patients seemed to have anxiety about expressing all aspects of the feelings involved: positive, anxious, and frustrated. The patient might anticipate rejection of his positive feelings, ridicule of his anxious feelings, and retaliation for his hostile feelings of frustration. The results can be interpreted in terms of Miller's theory of approach-avoidance conflict (14). In this connection, the irregularity in the increase of the curves in Figure 6 may represent conflict oscillation. The patients appear to overextend themselves in their expressions of feelings about the therapist and then, as anxiety about this increases, they withdraw and talk about something else. This is akin to the negative therapeutic effect (3).

The fact that the activity of the therapist has an effect on the increase of the patient's expression of feelings about the therapist may mean several things. First, the therapist's active remarks may be a way of indicating his interest in the area. Second, the active remarks of the therapist convey more information about the therapist's emotional reactions to what the patient is talking about than do the passive remarks. Since the emotional reactions of the therapist are primarily sympathetic and permissive, an increase in their communication should be encouraging to the patient to express his feelings.

A study of the relationship between the amount, quality, and rate of increase of therapist-relevant verbal material and the achievement of therapeutic success is immediately suggested by the present findings. However, no generally accepted criteria for success were readily available for the present sample.

V. Summary and Conclusions

The purpose of this paper was to describe the development, reliability, and validity of a method of studying psychotherapy. The method was a content analysis of the verbal behavior of the patient and the therapist. The categories of the content analysis were defined in terms of motivation and conflict, influenced by psychoanalytic and learning theories, and formulated in such a way as to be most relevant to an eventual understanding of the underlying emotional processes of psychotherapy.

A unit of patient verbal behavior was defined and shown to be reliable. A set of categories for the content of the patient's speech was developed from simpler categories used in several preliminary studies and shown to be reliable. The categories included sex, affection, dependence, and independence drives. Each drive was divided into positive approach, anxiety, and frustration. Additional categories were also used.

A set of categories for the remarks of the therapist was also developed and shown to be reliable.

The use of the method in a case study was illustrated by an analysis of the effects of subtle directiveness on the verbal behavior of a patient during the course of psychotherapy. It was found that the therapist mildly approved various subcategories of independence and mildly disapproved independence anxiety, intellectual defenses, and various subcategories of sex. As would be expected from learning theory, the approved patient-content categories increased and the disapproved patient-content categories decreased throughout the course of psychotherapy.

The use of the method in determining a general relationship was illustrated by a study of the patient's verbal expressions of feelings about the therapist. It was found that in a group of seven psychotherapy cases the frequency of statements about the therapist increased reliably ($p < .01$) as therapy progressed. Expression of all kinds of feelings tended to increase, rather than an increase in positive and a decrease in negative feelings. It was also found that the rate of increase of the frequency of patient statements about the therapist was positively related ($p < .01$) to the extent to which the therapist responds in an active rather than a passive way to patient material of all kinds, whether or not it is related to the therapist. However, it was also found that the therapists in the sample were significantly ($p < .01$) more active in responding to patient statements about the therapist than to patient statements not manifestly related to the therapist. These findings tend to fit in with what psychotherapists have previously known in a general, qualitative way.

There are some important limitations of the method. First of all, since it is based only on verbal behavior, some interpretation is needed to relate it to the underlying emotional processes. Second, in concentrating on motivation and conflicts, the method has ignored various cognitive processes such as insight. Third, the method is a laborious one and imposes limits on the size of samples that can be reasonably studied.

The validity of the method cannot be definitely established but only supported by its usefulness to date. The three case studies made with the aid of the method have all produced theoretically meaningful results. The first study (18) led to a hypothesis about displacement which was verified in the animal laboratory (21). The second study (20) yielded results which were in essential agreement with a detailed clinical analysis made of the same case (4). The last case study (Section III) fits in with expectations of learning theory (3, 11) concerning the effects of approval and disapproval on human behavior. The general relationships reported in Section IV seem theoretically meaningful, fruitful of further hypotheses, and in general agreement with clinical experience. Therefore, it is concluded that the method appears to be of some use and to have some validity.

Appendix A

MINOR STUDIES ON SUBTLE DIRECTIVENESS IN PSYCHOTHERAPY

I. *A group judgment of selected therapist remarks as approval and disapproval in the case of Herbert Bryan*

A special study was done in order to determine the degree to which a group of psychologically sophisticated individuals would agree on the approval or disapproval value of selected therapist remarks which seemed clear-cut to the writer. Since this study is essentially an experimental situation involving stimuli of hypothesized value rather than a reliability study of a content-analysis system, it was not presented in the main body of the study which is devoted to the content analysis proper.

The writer went through the case of Herbert Bryan (See Section III) and selected the 19 instances of *approval* which seemed most clear to him. He then selected clear instances of *disapproval,* which totaled 15. The 34 responses were marked off with enough patient context to make them understandable and then mimeographed in a random order in a little booklet. Twenty senior graduate students and young Ph.D.'s in clinical psychology were asked to judge what the therapist said as *approving, disapproving,* or *neither* in terms of the following definitions:

Approving: Any approval or agreement mixed with approval for: attitudes held by the patient; decisions to adopt attitudes; ways of behaving or talking in the therapy; past, present or future actions outside of the therapy.

Disapproving: Any slight disapproval, disagreement, criticism or argument with: the patient's attitudes; decisions to adopt an attitude; ways of behaving or talking in therapy; past, present or future actions outside of therapy. Also any indication that the therapist doesn't like the way the patient is discussing his problems or the language he uses in discussing his problems. Any time the therapist disagrees with the way in which the patient sees his problems and either offers a new definition of what's wrong or adds to the patient's interpretation. (This is to be scored "D" even if the scorer tends to agree with the therapist's view rather than the patient's.)

Neither: Anything to which the approving or disapproving do not apply.

The results are shown in Table 15. The group reliably judged 15 of the 19 approvals, and 10 of the 15 disapprovals, in the correct categories.

TABLE 15

THE AGREEMENT IN JUDGING 34 SELECTED THERAPIST'S REMARKS AS APPROVAL AND DISAPPROVAL BY TWENTY CLINICAL PSYCHOLOGISTS

Judgment of experimenter (Murray)	Judgments of 20 Clinical Psychologists			
	Judged approval*	Judged disapproval*	Judged neither*	No agreement among judges
Approvals	15	0	0	4
Disapprovals	0	10	1	4

* The number in the body of the table for the first three columns is the number of therapist's remarks that a reliably large part of the group ($p < .02$) judged as approval, disapproval, or neither.

No approval was mistaken for a disapproval, and no disapproval was judged as approval.

The results of this study tend to support the finding in Section III that the case of Herbert Bryan contains remarks by the therapist that have mild approval or disapproval value. The study, of course, offers no further support for the other findings in Section III that the approval and disapproval formed a pattern, and were related to changes in the verbal behavior of the patient.

II. *An opinion study of the subtle directiveness in one psychotherapy hour in the case of Herbert Bryan*

The data presented in Section III support the view that the therapist influenced the patient's decision to be independent. However, the judgments of short segments of the therapist's behavior may have produced a distorted picture. After all, the patient responds to a total impression. In addition, the judgments of clinical psychologists may not be comparable to those of Mr. Bryan. Therefore, a general impression of the directiveness of one hour was gotten from a heterogeneous group of 20 intelligent but relatively naive people. The people were undergraduates at Yale College and New Haven State Teachers College, relatively untrained Army psychology technicians, and Yale graduate students. The hour chosen was the sixth, the one just before the major change. The people were asked to read a mimeographed copy of Hour 6 and answer four simple questions.

The first question was as follows: "In this psychotherapy hour the patient is discussing a strong conflict. He is trying to make a difficult choice. This choice is between a)———and b) ———." The purpose of this question was to see if the subject understood the basic conflict in the hour. All subjects answered quite adequately in their own terms.

The second question was: "Do you feel the therapist basically: a) wants the patient to make one choice rather than another, or b) is completely neutral and doesn't care which choice the patient makes?" The purpose of this question and the next was to allow the subject to answer that he felt the therapist had an opinion, possibly just because he was a therapist, but still to answer "no" to a later question about the therapist acting on his opinion. Nineteen of the 20 subjects answered that the therapist basically wanted the patient to make one choice rather than another ($p < .0001$).

The third question was: "If you feel the therapist wants the patient to make a choice, which direction do you think the therapist wants the patient to take?" All of the 19 people left after question 3 gave an answer which indicated, in their own words, that the therapist wanted the patient to be independent ($p < .0001$). To some extent at least, the answers to the two questions above are probably due to the expectations that people have of therapists. There is ample evidence that Mr. Bryan also had these expectations. However, the answers to the next question will show that Mr. Bryan was not influenced by his expectations alone.

Question 4 reads as follows: "Do you feel that the therapist gets this across: a) very obviously and strongly, b) in a subtle and indirect way, or c) doesn't get it across at all." Two people felt that the therapist got it across very obviously and strongly, 16 people felt he got it across in a subtle and indirect way, and only one person felt he didn't get it across. Therefore, the therapist got across his preference for independence ($p < .0001$) and did it in a subtle way ($p < .01$) according to the opinions of this group of people.

The results tend to support the finding in Section III that the therapist employed subtle directive techniques which influenced the verbal behavior of the patient. The agreement between the common sense judgments in this study and the results of the content analysis in Section III lend some support to the validity of the content-analysis system.

APPENDIX B

INSTRUCTIONS FOR SCORING PATIENT UNITS

Simple sentences. The simple sentence is the basic unit of scoring. In its purest form the sentence contains a subject and predicate—the predicate usually contains a verb and an object. Adjectives and adverbs may be present. The simple sentence may not make complete sense all by itself since the meaning of pronouns, allusions, etc., may gain their meaning from the context. But with this help the simple sentence must contain a complete idea. Some examples are:

a. "The telephone wires were cut."
b. "It's been in my conscience."
c. "But I couldn't stand it."
d. "I'm terribly upset about this experience last night."

Incomplete sentences. I. A statement may be incomplete in itself and yet not be part of a previous or following statement. If the missing part is strongly implied then it is counted as a regular unit, e.g.,

a. "Parked in the driveway." (Here it is implied that "The car was parked in the driveway.")
b. "A young bride who is madly in love with her husband." (Here it is implied that, "She is a young bride. . . .")
c. "Very exciting over there." ("It was very. . . ." is implied.)

II. But where the phrase is patently incomplete and never followed up it is ignored, e.g.,

a. "And . . . ah . . . they to . . . she, etc." (Here the "they to" leads to nothing.)
b. "It must of . . . I . . . I can't, etc." (Here the "It must of" leads to nothing.)

Slightly complex sentences. These are sentences with introductory phrases, dependent clauses, adjectival phrases, immediately preceding or following explicatives, etc. Some examples are:

a. "I lied to you last week, last Tuesday."
b. "I've been very dishonorable, many times."
c. ". . . or how you'd feel about all this—these things I told you."
d. "But I still have that terrible guilt feeling that I can't do these things."
e. "I don't remember anything about it—only that I had this nightmare."
f. ". . . a young bride of three months who is madly in love with her husband."
g. "I said it happens but it isn't a good idea to follow through on it."
h. "A close friend of mine, a very close friend, just died."
i. "Oh! Oh! But its, etc."
j. "God! I can't believe it."

Relational sentences. These are sentences containing two parts which are related in some way such as showing cause and effect, explanation, definitions, etc. A simple example is: "I just had to tell you that I lied because I'm not honorable at all." The cause and effect may be logical or not and may be vague or specific. These sentences are scored as one unit. The commonest words used in these sentences are: because, since, if . . . then, so, in spite of, therefore, when, before, after, etc. However, these words are only guides since they may be used in other ways and other words may be used to relate the two parts. Furthermore, the relation may be implied (strongly) without a specific relational word; e.g., "I mean I figured Alice is growing up and she needs a respectable mother."

Conjunctival sentences. These are sentences which have several phrases joined together such as: "Well he was kind and sweet." They usually contain a conjunction such as: and, or, either . . . or, etc. But two types can be distinguished. In the first type a number of nouns, or verbs, or objects may be joined but the rest of the sentence is unitary; e.g.,

a. "Mother and Father treated me like a little girl."
b. "I don't know what's going to become of my husband and children."
c. "She was always nagging and shouting at me."

These sentences are scored as one unit. On the other hand the conjunction may combine two sentences complete or nearly complete in themselves. There may be separate ideas or the same idea repeated. These sentences are scored as two or more units. However, a verb and a subject must be present in each section of the sentence to be scored as a separate unit; e.g.,

a. "I didn't know what you'd think/or what you'd say/or how you'd feel about all this . . . these things I told you/." (In this case three units share the same final ending.)
b. "We went visiting the other afternoon/and I had a couple of drinks/."
c. "Bud is sweet/and he's kind/and he has a good disposition/."
d. "But it never comes up/or does it?/"

Sentences related to the therapist's remarks. I. A thought unit may be interrupted by the therapist but finished afterward. The stroke is put at the end of the continuation; e.g.,

P: "- - - -/They took her car as a get-away. . . ."
T: "Oh, I see. MmmHmm."
P: " . . . car./"

However, the continuation must be a definite part of the first line. If a new thought is introduced, if the first thought is repeated, or if the first thought is modified with respect to what the therapist says, then it is scored separately. If in this case the first line is left incomplete it is ignored.

II. Thought units which are incomplete in themselves but which serve as agreement, disagreement, etc., with a therapist statement are counted as units; e.g.,

a. T: "I think that it's very important that you understand why you weren't able to tell me that last week. You see?"
 P: "Mmmm/Oh, I was afraid to.? I. . . ."
b. T: "That's what you were dreaming."
 P: "Evidently./"
c. T: "Mmmhmm. And then what happened?"
 P: "I dunno/It's a long. . . ."
d. T: "But you know he isn't. . . ."
 P: "Yes, but / still I can't. . . ."
 (This is a special case. Wherever there is a "Yes, but," "Still," "Yes, however," etc., reaction after a therapist remark it is scored as a unit.)

III. Sometimes the patient will finish a sentence for the therapist. In this case the whole idea is scored as a patient unit; e.g.,

a. T: "- - - - Understanding it now. . . ."
 P: "Will make a change./"
b. T: "- - - - for something which just. . . ."
 P: "happened./"

Questions and quotes. All questions are to be scored as a unit even if they are tacked on the end of a sentence; e.g.,

a. "Because its gone on too long./ Hasn't it?/"

The introduction to a question or quote is scored along with it and not separately;

b. "And I said 'I don't know'."/

Two sentences within a quote are to be scored separately as other sentences are;

c. "And I said 'I don't know.'/ 'I just have a guilt feeling'."/

Appendix C

INSTRUCTIONS FOR SCORING PATIENT CONTENT

Disturbances of free association.

a. Asking questions, seeking directions, entering into bilateral discussion with therapist; e.g., "Should I just try to remember my dreams or jot them down?" "What do you think I should talk about now?"
b. Not following the rule of free association; enlisting therapist in this; e.g., "I can't talk about my father." "You don't want me to relate trivial matters, do you?" "That's sort of personal."
c. Expressing difficulty in talking or thinking of things to say; e.g., "My mind is a complete blank." "We've talked about everything now."
d. Other.

Agreement with therapist remarks.

a. "Mmmhumm," "yes," "surely," etc.
b. "I agree," "I think so," etc.
c. "No"—if the therapist's remarks required "no" for agreement; e.g., T: "But you don't want that!" P: "No."
d. Other.

Disagreement with therapist remarks.

a. "No, I disagree," "I don't think so," etc.
b. "Yes, but . . . ;" "I agree, but . . . ;" "that's so, but . . . ;" etc.
c. "Yes," if the therapist's remarks required "No" for agreement; e.g., T: "But you don't want that!" P: "Yes, I do."
d. Alarm or startle reactions: "Oh my!" "Goodness!"
e. "I don't know." "I'm not sure," etc.
f. Other.

Intellectual discussion. Includes discussion of abstract psychological theory, philosophy, art, literature, science, politics, etc. (except where the activity is a professional one.)

a. "Psychologically and philosophically one could attribute it to conditioning."
b. "I am fond of the Existential outlook."
c. "I'm writing a philosophical novel."
d. "Our values are our private absolutes."
e. "Human behavior is subject to variation anyway."
f. "There are two kinds of will."
g. "Toynbee discusses that problem rather well."
h. "But the cultural anthropologists have seriously undermined Freudian theory."
i. "I must tell you about the concert I went to last night."
j. "I'm trying to reconcile scientific determinism with my own inner feeling of free will."

Generalized anxiety. Includes all psychological and somatic expressions of anxiety which are not related to a drive nor related to any specific person or object; general "free floating" anxiety and guilt.

a. "I feel panicky about the thought of death."
b. "I tremble and then it would ease off, then I'd start again . . . in waves."
c. "I feel tense as if there's some force inside of me trying to get out."

d. "I have frightening suicidal thoughts."
e. "Life is meaningless . . . why are we here?"
f. "I feel like there's an axe pressing down on my dynamo."
g. "I feel like a coward."
h. "I don't know how I've lived so long and learned so little."
i. "I dream of monsters trying to catch me and eat me up."

Sex. Includes all statements referring to the positive or approach component of the sexual drive; direct expression of sexual needs and wishes, description of sexual attraction and arousal, sexual activity not mixed with fear or guilt, planning for sexual satisfaction, courtship and dating among unmarried people where the erotic element is present but institutionalized, description of homosexual feelings and other perversions, descriptions of masturbation, discussion of normal sex education.

a. "I talked over my affair with my father."
b. "He wanted to have sexual relations with me."
c. "I've had funny feelings in my legs since coming to you."
d. "I used to like to cling to my brother."
e. "I have dreams about having intercourse with a faceless woman."
f. "I liked Jewish men because I thought they were sensuous."
g. "I menstruated freely and easily."
h. "I like to watch strip teasers."
i. "I look on myself as a sexual artist."
j. "I value sex highly."

Sex anxiety. Includes expressions of fear, anxiety, or guilt about sex; nervousness, irrational fears, phobias, compulsions, depression, hopelessness, confusion, helplessness, conflict, tension, blocking; avoidance behavior, denial of a sex drive, negative attitude toward sex, rationalizing one's behavior, deprecating self; feelings of inadequacy, somatic symptoms, impotence, frigidity.

a. "I'm frightened by my sexual feelings toward you."
b. "I could only get aroused during intercourse by pretending I was completely paralyzed."
c. "I was nervous and disgusted when my brother made advances to me."
d. "I felt awful when I performed fellatio on him."
e. "I was afraid someone would discover us."
f. "I only felt aroused when I found out he was impotent."
g. "I worry about functioning adequately."
h. "My diaphragm is the next to the smallest size and I felt I was underdeveloped."
i. "My breasts didn't grow until after all my girl friends had fully developed."
j. "I nearly threw up when I started having intercourse with her."

Sex frustration. Includes hate, anger, resentment, criticism, or complaint about the frustration of the sex drive; description of behavior designed to, or expression of desire to, destroy, hurt, attack verbally or physically, disparage, humiliate, domineer, or deprecate in any way; frustrating another person, withholding the symbols of affection, making someone uncomfortable.

a. "My husband has premature ejaculation."
b. "He's mechanical . . . no tenderness."
c. "My mother punished me for talking about sex to a friend of mine when I was a boy."
d. "On our wedding night he sat up all night talking to the cab driver."
e. "He was constantly experimenting."
f. "You won't do anything here in therapy."
g. "My mother wouldn't let me feel proud of my first menstruation."
h. "He forced me to do fellatio."
i. "I have intense feelings of jealousy."
j. "Intercourse just isn't readily available in the environment."

Affection. Includes direct expression of needs for love and affectionate behavior; needs for and appreciation of affection, friendship, acceptance, understanding; feeling of loving, giving affection, etc., or having entered into an affectionate relationship; doing something positive for or with a person; holding someone in high regard, respecting, appreciating etc.

a. "I like him, he's such a fine person."
b. "I like to please my mother."
c. "I have a euphoria before my interview."
d. "My little son helped me on the farm today."
e. "I went out with him because I could talk to him the way I used to with my father."
f. "I feel at ease talking to you."
g. "I want to settle down and marry."
h. "I'm lonely . . . I want someone to love."
i. "He's intelligent, good-looking, and kind."
j. "She's a wonderful gal."

Affection anxiety. As in sexual anxiety but related to affection.

a. "I'm afraid to fall in love."
b. "I don't want to become emotionally involved with a girl until I finish the treatment."
c. "I'm usually not close to my son."
d. "After that experience I don't feel like dating any more."
e. "I never did love my husband."
f. "I get upset and feel nauseous when I try to talk to people at a party."
g. "I'm going to stop my husband from trying to see me."
h. "I'm going to stop dating her."
i. "A really decent man wouldn't find me very interesting."
j. "You don't love me because I'm so stupid and neurotic."

Affection frustration. As in sex frustration but related to affection.

a. "My parents never gave me any love or attention."
b. "I want to marry you and you keep telling me about other men."
c. "I always felt my brother was the favored one."
d. "Women just use love to sugar-coat their sexual feeling."
e. "My wife wouldn't stay with me during the vacation."
f. "We didn't feel accepted socially."
g. "He insisted on having sex while that poor dog of mine lay next to the bed dying."
h. "She won't show me any love."
i. "He never gave me any presents."
j. "There aren't any eligible men around."

Dependence. Includes expressions of needs to depend on someone, let someone else take the initiative, to be told what to do, to be helped, to be cured by an outside agent; description of dependent behavior or nurturant actions by others to self; being appreciative of being taken care of by another; making personal security contingent on another.

a. "I liked being cared for by my father when I was so sick."
b. "I have a dentist I like and trust."
c. "I like to have someone I can confide in and get advice from."
d. "I'm hoping that after I talk myself out you're going to pull a rabbit out of the hat."
e. "I need someone to help me out of a bad situation."
f. "Grandmother used to send me money."
g. "I want to marry a man who is superior to me."
h. "I'm looking for a magic key to my problems."
i. "You are my superior in emotion changing technique."
j. "I need help."

Dependence anxiety. As in sex anxiety but related to dependence.

a. "I'm afraid of people stronger than me."
b. "I don't like to accept any of my husband's money."
c. "I'm afraid of being in a situation when I can't move—when I'm at the mercy of other people."
d. "I know that people can make me do things if they're smart enough."
e. "I was afraid to start therapy."
f. "I get nervous when I date a really strong man."
g. "I don't trust dentists."
h. "I don't like to accept charity."
i. "I would be frightened by a domineering wife."
j. "I couldn't let on that I needed her."

Dependence frustration. As in sex frustration but related to dependence.

a. "It seems that only lonely, clinging people like my husband love me."
b. "My mother wouldn't send me the money when I needed it."
c. "You keep throwing me on my own."
d. "She won't help me with the work."
e. "My parents wouldn't send me to school."
f. "I need to be taken care of and he's too weak to do it."
g. "You can't confide in anyone in the family —they're all independent."
h. "She used to tell me all her problems but I wanted to tell her mine."
i. "I'm discouraged—I don't think this is doing me any good.
j. "My wife just doesn't do things for me the way my mother used to."

Independence and self-assertion. Includes expression of needs to be independent, adult, mature, assertive, ambitious, competitive; standing up for one's rights, defending against attack, saying "no" when one has to, expressing an opinion or warranted criticism; descriptions of behavior or plans to be independent, etc.

a. "I pick out my own clothes now."
b. "I have more ambition lately."
c. "I love work that requires thinking and responsibility."
d. "I'm going to arrange it so that mother can't move in with me."
e. "I told him that's how much work I did and that he'd have to pay for it."
f. "I'm successful at everything when I'm feeling well."
g. "You can only achieve growth by meeting new situations."
h. "I can relax when I'm playing cards with her now."
i. "I have to accept responsibility for my neurosis."
j. "I enjoy getting people to cooperate with me."

Independence anxiety. As in sex anxiety but related to independence.

a. "I used to be afraid to assert myself to my mother."
b. "This new job is more uncertain than the old."
c. "I'm blocked in every field of effort."
d. "I feel guilty about doing something that I like to do—like building ship models."
e. "I get anxious when having a philosophical argument."
f. "My stomach gets upset when business is bad."
g. "I'm no good—I can't cope with a job."
h. "I felt I couldn't live up to their expectations."

i. "I'm just lazy."
j. "I get tired very easily."

Independence frustration. As in sex frustration but related to independence.

a. "She picked out all my clothes."
b. "I had to be home by 11 o'clock."
c. "He won't let other people talk—he keeps interrupting them."
d. "Jobs are hard for women to get."
e. "She reads my diary and mail."
f. "I don't like to be bossed."
g. "I don't like the fact that you run the hours—that you know more than I do."
h. "She controlled the whole situation and I didn't like being at her beck and call."
i. "He won't let me express my opinions."
j. "I hate my mother for the way she dominated us."

Unspecified anxiety. As in sex anxiety but related to no apparent drive. To be distinguished from generalized anxiety by the fact that it is related to some specific person or thing.

a. "My brother used to frighten me when I was a child."
b. "My husband does frightening things like once he cut off the head of my marionette."
c. "I wake up in a cold sweat after dreaming about my father."
d. "My heart beats fast and I tremble when I come in here."
e. "I don't feel worthy of taking up your time."

Unspecified frustration. As in sex frustration but related to no apparent drive.

a. "I just hate her."
b. "I noticed what a big nose you have."
c. "I get irritated every time I see my brother."
d. "He was a clumsy dancer."
e. "When he came home drunk I hit him on the head."

APPENDIX D

INSTRUCTIONS FOR SCORING THERAPIST CONTENT

Instruction to free-associate. Whenever the therapist explains the rule of free association, the need for it, the benefits to be gained from following it, or directly insists on it. Whenever the therapist asks the patient to tell him what he was thinking about during a pause.

a. "You must speak out your thoughts."
b. "Even if it disturbs you, you must strive to say anything that comes into your mind."
c. "If you say everything that comes into your mind we can find out what's bothering you and help you."
d. "What were you just thinking about?"

Labels. Wherever the therapist tells the patient what it was he was feeling or doing. Wherever the therapist points to the patient's own feelings or behavior in a situation where the patient has only talked about the other person. Most causal statements are labels.

a. "You must have felt *excited* when he did that."
b. "That must have been a *frightening* experience."
c. "Or perhaps you feel that way whenever you have to *depend* on someone."
d. "Couldn't it be—that it was a matter of your letting her do it, rather than her doing it?"

In scoring labeling remarks by the therapist, the symbol Lab. should be followed by the general nature of the labeling. The examples above would be scored:

Lab.—sexual excitement—brother
Lab.—fear—automobile crash
Lab.—dependency—general
Lab.—rejecting daughter

Part of the general nature of the label is found in the patient content, e.g., the fact that "he" in the first example refers to the patient's brother.

Discrimination. Wherever the therapist points out a difference between two situations, feelings or behavior patterns. The main instances are where he points out the difference between past and present conditions, and the difference between what the patient expects from people and what they are like.

a. "When you were a child your mother didn't let you think for yourself but now you're an adult and you're expected to be independent and think for yourself."
b. "Well, not all men are like that."
c. "Most people don't find sex as nauseating as you do."

Again the two things discriminated are written next to the symbol. For those above:

Disc.—child and adult—independence
Disc.—good and bad men—selfish
Disc.—patient and other people—sex as nauseating

Similarities. Wherever the therapist points out two situations, feelings, etc., which are similar or have a common element. It may be done in the form of a question or leading statement.

a. "That's the same way you felt about your first husband."
b. "Isn't that the same way you felt when you were a child?"
c. "Isn't this the way you react to all men?"

Again the two situations which are related are noted. Above:

Sim.—First and second husbands

Sim.—Child and adult
Sim.—Reaction to all men

Strong approvals. Any form of strong approval for the way the patient has just been speaking or for an outside action that the patient has described. Strong agreement, encouragement, reassurance, and explicit permissiveness.

a. "That sounds excellent."
b. "I appreciate how tough it was for you to say that."
c. "You're certainly making progress now."

Disapprovals. Means some indication that the therapist doesn't want the patient to continue talking about what he has just been talking about, or the way in which he was talking about it. Any indication that the therapist does not feel the topic is important, interesting, or acceptable. Any surprise, alarm, anxiety, or disturbance on the part of the therapist. Any of the typical social punishments such as: belittling the patient or his statements: holding up an example in contrast to what the patient says; replacing a patient's interpretation with the therapist's own and different one in such a way as to reject what the patient has said. Giving a negative evaluation to something the patient says.

a. "Yes, but this may not be the most important problem facing you."
b. "You talk on such an abstract level that it's difficult to understand what it means to you."
c. "Do you really do *that*?"
d. "You really can't mean that."
e. "We aways get back to your mother."
f. "That may be, but don't you think it's rather your hostile feelings?"
g. "That's not getting you any place."
h. "You can't do both."

Demands. Is an attempt to elicit verbalization about a topic that hasn't immediately preceded the demand. That is, a unit is scored as a demand only if it is an attempt to bring up a new topic or one which has been dropped by the patient. In order to qualify as "new" a topic must not have been mentioned by the patient in the ten patient statements just previous to the therapist's remark.

a. "You were telling me earlier about your mother, what about her?"
b. "I wonder if you would tell me something about your sex life."
c. "I was thinking about how you went about getting a job last month. How did you feel about that?"

Directions. This is wherever the therapist tells the patient what to do (other than to free-associate). Also any statements of opinion about general questions.

a. "Perhaps if you tried to be nice to him it would help."
b. "Don't you think you should keep trying?"
c. "I think there are greater satisfactions in marriage."

Mild probes. This category is for instances where the therapist is encouraging what the patient is talking about but where he asks the patient to continue talking about the topic, asks for elaboration of one point, or asks for information directly related to what the patient is talking about. It will also include confession of ignorance or lack of understanding on the part of the therapist. The typical "tell me more" remark.

a. "That seems strange, doesn't it?"
b. "How could that be?"
c. "Who would she belong to then?"
d. "What do you mean by 'passion'?"
e. "Would you tell me a little more about that?"
f. "I don't completely follow that."
g. "Why did you do that?"
h. "It sounds like there's more of a story behind that."

Mild approvals. This category is for instances where the therapist shows a minimum response which is generally positive.

a. "Yes."
b. "Oh, I see."
c. "I understand."

"Mm." This category is for instances when the therapist grunts or verbally nods. It includes "Mm" and "MmMm" but includes no words.

Irrelevant. All general opening and closing remarks; discussion of technical details in therapy such as "O.K. Tuesday at 10"; other irrelevant remarks. In addition, any response on the part of the therapist which cannot be put in other categories. Any unscorable self-phrases or mumbling.

BIBLIOGRAPHY

1. Auld, F., & Murray, E. J. Content-analysis studies of psychotherapy. *Psychol. Bull.,* 1955, **52,** 377-395.
2. Dollard, J., & Mowrer, O. H. A method of measuring tension in written documents. *J. abnorm. soc. Psychol.,* 1947, **42,** 3-32.
3. Dollard, J., & Miller, N. E. *Personality and psychotherapy.* New York: McGraw-Hill, 1950.
4. Dollard, J., Auld, F., & White, Alice M. *Steps in psychotherapy.* New York: Macmillan, 1953.
5. Fenichel, O. *The psychoanalytic theory of neurosis.* New York: Norton, 1945.
6. Freud, S. *Collected papers,* Vol. III, London: Hogarth, 1950.
7. Freud, S. *A general introduction to psychoanalysis.* Garden City, N.Y.: Garden City, 1943.
8. Fromm-Reichmann, Frieda. *Principles of intensive psychotherapy.* Chicago: Univer. of Chicago Press, 1950.
9. Greenspoon, J. The effect of a verbal stimulus as a reinforcement. *Proc. Ind. Acad. Sci.,* 1950, **59,** 287.
10. Grummon, D. L. An investigation into the use of grammatical and psycho-grammatical categories of language for the study of personality and psychotherapy. Unpublished doctor's dissertation, Univer. of Chicago, 1950.
11. Hull, C. L. *Principles of behavior.* New York: Appleton-Century-Crofts, 1943.
12. Janis, I. L. Meaning and the study of symbolic behavior. *Psychiatry,* 1943, **6,** 425-439.
13. Lasswell, H. D. Certain prognostic changes during trial (psychoanalytic) interviews. *Psychoanal. Rev.,* 1936, **23,** 241-247.
14. Miller, N. E. Experimental studies of conflict. In J. McV. Hunt (Ed.) *Personality and the behavior disorders.* New York: Ronald, 1944. Pp. 431-465.
15. Miller, N. E. Theory and experiment relating psychoanalytic displacement to stimulus response generalization. *J. abnorm, soc. Psychol.,* 1948, **43,** 155-178.
16. Miller, N. E., & Kraeling, Doris. Displacement: greater generalization of approach than avoidance in a generalized approach-avoidance conflict. *J. exp. Psychol.,* 1952, **43,** 217-221.
17. Miller, N. E., & Murray, E. J. Displacement and conflict: learnable drive as a basis for the steeper gradient of avoidance than of approach. *J. exp. Psychol.,* 1952, **43,** 227-231.
18. Murray, E. J. A case study in a behavioral analysis of psychotherapy. *J. abnorm. soc. Psychol.,* 1954, **49,** 305-310.
19. Murray, E. J., & Miller, N. E. Displacement: steeper gradient of generalization of avoidance than approach with age of habit controlled. *J. exp. Psychol.,* 1952, **43,** 222-226.
20. Murray, E. J., Auld, F., & White, Alice M. A psychotherapy case showing progress but no decrease in the discomfort relief quotient. *J. consult. Psychol.,* 1954, **18,** 349-353.
21. Murray, E. J., & Berkun, M. M. Displacement as a function of conflict. *J. abnorm. soc. Psychol.,* 1955, **51,** 47-56.
22. Rogers, C. *Counseling and psychotherapy.* Boston: Houghton Mifflin, 1942.
23. Rogers, C. *Client-centered therapy.* Boston: Hougton Mifflin, 1951.
24. Seeman, J. A study of the process of nondirective therapy. *J. consult. Psychol.,* 1949, **13,** 157-168.
25. Snyder, W. U. (Ed.) *Casebook of nondirective counseling.* Boston: Houghton Mifflin, 1947.

(Accepted for publication November 9, 1955)

Vol. 70, No. 14 Whole No. 421, 1956

Psychological Monographs: General and Applied

Sex-Role Preference in Young Children[1]

DANIEL G. BROWN[2]
Neuropsychiatric Service, 3275th USAF Hospital,
Parks Air Force Base, California

INTRODUCTION

SEX-ROLE behavior is one of the least explored areas of personality formation and development. Although there have been a number of techniques developed for the study of masculinity-femininity patterns in adults (13), comprehensive reviews of the literature on sex differences (3, 26) reveal a relative lack of data in this area as far as children are concerned. While the physical, mental, and various social aspects of child development have been increasingly investigated, especially in the last few decades, the specific problem of sex-role behavior has been largely neglected.

Despite the poverty of research in this area, the importance of sex-role adjustment is commonly recognized. It is one of the integral components of normal, satisfactory social development and adjustment. Mowrer has summed up its importance with these words: "Personal normality presupposes that an individual has assimilated not only those values and ideals which are regarded as necessary and proper for all persons, but also those values and ideals which are uniquely appropriate to one's *sex role,* as a man or as a woman" (16, p. 615). From the clinical fields of psychology and psychiatry, too, have come increasing recognition of the relationship between disturbances in the individual's sex-role functioning and personality maladjustments. In fact, failure to establish appropriate sex-role identification is hypothesized as a predisposing factor in personal maladjustments including perversion, neurosis, and psychosis. Mowrer has particularly stressed this view in his discussion of sex-role identification and neurosis: "Other writers have not failed to comment upon the relationship between identification difficulties and the predisposition to neurosis; but they have not, the author believes, sufficiently stressed the importance of this connection or given a proper indication of its generality" (16, p. 596). In short, then, the area of sex-role behavior constitutes a vitally important and vastly significant problem.

LEARNING AND ENVIRONMENTAL FACTORS IN SEX-ROLE BEHAVIOR

Although biological differentiation of the sexes is an innate process, the be-

---

[1] This monograph, part of which was presented at the 1954 meeting of the American Association for the Advancement of Science in Berkeley, is based upon a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Psychology, Graduate College, of the University of Denver. Acknowledgment with gratitude is made to Dr. E. Ellis Graham for direction of the investigation and to Dr. Alfred B. Shaklee for his valuable advice and encouragement. For helpful advice and suggestions concerning the present monograph, the author is indebted to Dr. O. Hobart Mowrer, Dr. Robert R. Sears, Dr. Lewis M. Terman, Dr. Harold E. Jones, and Dr. Georgene H. Seward.

[2] On leave of absence from the Department of Psychology, University of North Dakota.

Bureau Ednl. Psy. Research
DAVID HA [illegible] ING COLLEGE
Dated . ...
No. ....

havior patterns, including overt actions, interests, values, and preferences, associated with such dichotomous biological differentiation appear to be governed in large measure by social learning conditions. Thus, Seward observes: ". . . the individual is trained to his sex role from the moment of birth when girls are placed in pink, boys in blue bassinets" (21, p. 153). And Ferguson (10) in a study of masculinity-feminity patterns in college students found that childhood learning experiences were crucial determinants of adult sex-role behavior and adjustment.

From the above discussion it would seem to follow that whether a person is male or female *biologically* is dependent upon genetic and biological processes, but whether that person is male or female *socially* and *psychologically* (i.e., in terms of sex-role behavior) is in large measure dependent upon learning, environmental factors, and experiential development.

Hypotheses and Implications

The relevant hypotheses and implications of the present investigation may be briefly summarized in terms of four considerations: (*a*) the nature of identification, (*b*) sex differences in the acceptance of sex roles, (*c*) culturally determined conflicts in sex-role behavior, and (*d*) homosexuality in relation to sex-role development.

The nature of identification. As defined in the present study, identification is a process involving a child's learning to think, feel, act and become like parental figures and other significant persons and groups in his life. Relative to the identification process and parent-child relations, psychoanalytic writers, beginning with Freud, have distinguished between identification-object choice, where the parent is what the child wants to be, to become, to be like; and love-object choice, where the parent is what the child wants to possess as a love object.

Mowrer (16) has discussed the process of identification within the framework of learning theory and has postulated three broad outcomes and three basic personality patterns: *(a)* identification with the same-sex parent results in *normal* adult sex-role behavior (normality), *(b)* identification with the opposite-sex parent or model results in *inverted* adult sex-role behavior (inversion and perversion), and *(c)* identification of a confused, ambivalent nature with both parents or sex models results in *neurotic* adult sex-role behavior (neurosis or maladjustment). The neurotic or maladjustive pattern presumably involves a bisexual type of identification in adulthood in which the individual would be mixed up or confused relative to the masculine and feminine components in his personality.

In connection with the above discussion, several significant problems arise out of Mowrer's formulations: First, can sex-role preference patterns in young children be demonstrated and quantified? Second, are such preferences maintained in later childhood, adolescence and maturity? Third, what social and psychological factors are functionally related to differing sex-role preferences? The present investigation, which has its inception in Mowrer's provocative thinking in this area, is in part an attempt to answer the first of these questions; if this can be done satisfactorily, attacks on the second and third problems will be facilitated.

Sex differences in the acceptance of sex roles. Girls, according to Freud, and other psychoanalytic writers, inevitably experience greater difficulty in accepting their sex roles than do boys. The reason for this is to be found in the *anatomical difference* between the boy and girl; the little girl allegedly experiences the shock of her life when she discovers she is without the genital organ possessed by the little boy. This is supposed to arouse intense resentment and envy in every girl, drives her in one way or another in the direction of masculinity, and, hence, makes acceptance of the feminine role difficult.

In contrast to the emphasis Freud places on the anatomical difference between the boy and girl, Adler (1) and other personality theorists have emphasized the greater *sociocultural advantages* that the male enjoys compared to the female. This is assumed to arouse envy, breed contempt, result in feelings of hostility, and make it difficult for the little girl to want to be a little girl. She comes to perceive varying degrees of futility in the female role, since it becomes almost synonymous with inferiority, weakness, and lowliness, whereas the little boy early becomes aware of the virtue of masculinity and its equation with superiority, independence, and loftiness. This state of affairs results in what Adler calls a *masculine protest* that is manifested in girls by masculine role preference and in boys by an exaggeration or compulsive clinging to masculinity.

It would seem to follow from the above discussion that both Freudian and Adlerian theory, though on very different bases, assume that the average girl experiences greater dissatisfaction in accepting the feminine role than is true for the average boy in accepting the masculine role. From these formulations a straightforward hypothesis emerges, namely, whether because the male has something biological that the girl lacks, or because the male has greater cultural privileges not granted to the girl, or because of both these advantages, an investigation of sex-role preference in children will show that girls tend to prefer the masculine over the feminine role, or at least will show a significantly greater trend toward masculinity than boys show toward femininity.

A third consideration is that of *culturally determined conflicts in sex-role behavior.* Some writers assume conflicts in sex-role adjustment to be more frequent and pronounced in girls than in boys because of such factors as: *(a)* the incompatibility arising from the fact that the girl is biologically female but aspires psychologically toward masculinity; *(b)* career alternatives faced by the girl, i.e., vocational, homemaker, or both; *(c)* insufficient or irregular encouragement and recognition for feminine tendencies in childhood, in that the "tomboy" pattern may bring as much approval as the "little lady" pattern; and *(d)* the modern status of females today compared to former generations; in the past the girl perceived clearly her status and there was a minimum of ambiguity and uncertainty concerning her role, but today the feminine role is no longer clear-cut and singular, and the result is conflict.

Such factors as the above give rise to another hypothesis relative to the sex-role development of children: girls will tend to show greater ambivalence or confusion in sex-role preference than will boys.

In contrast to the above formulation, other writers have emphasized the comparatively greater difficulty of boys in acquiring their sex role than is true for girls (17, 18, 12). The basis for this assumption is in the fact that boys experience less exposure to the appropriate identification model than is the case with girls. The father is typically absent from home much of the time and hence there is less opportunity for the boy to perceive and identify with the masculine role. The little girl, however, is able to understand and participate in many of the important aspects of the adult feminine role since the mother is typically around much of the time, especially in the earlier years. This is also true in the case of broken or separated homes, since a mother substitute is easier to provide than a father substitute (4). These considerations might be interpreted as leading to the hypothesis that boys will encounter *greater difficulty* in sex-role development than would be true for girls.

The above formulations point in different directions, i.e., one conception suggesting greater difficulty for girls, because of their dissatisfaction with the feminine role, and the other conception suggesting greater difficulty for boys because of their inadequate exposure to the masculine role model. In connection with these diverse predictions, Sears[3] has made an important conceptual contribution that clarifies the basis for this diversity. He has emphasized the need to differentiate between sex-role *identification* and sex-role *preference.* Most, if not all, writers, including the present author, have tended to use these terms synonymously. Sex-role *identification* refers to the behavior associated with one sex or the other that the individual introjects and acquires as his own, whereas sex-role preference refers to the behavior associated with one sex or the other that the individual would like to adopt, or that he perceives as the preferred or more desirable behavior. It is probably true that in most cases an individual may be said *both* to *identify* with a given sex model as well as to *prefer* that sex model, i.e., he identifies with what he prefers and prefers that with which he identifies—in such a case there is no inconsistency, no internal conflict, and the two processes would tend to fuse into a single phenomenon. Such consistency however, does not always prevail; thus, a person might *identify* with the appropriate sex model but nevertheless retain varying degrees of *preference* for the other sex model. With this distinction in mind, then, boys might be expected to show greater acceptance or liking for their role in terms of sex-role preference than would be true for girls; but in terms of sex-role identification, girls might be expected to have an easier course of development, identifying earlier and more thoroughly with the sex-appropriate role than would be true for the boys.

A final consideration is that of *homosexuality in relation to sex roles.* That there is a relationship between male sex-role inversion and *passive* male homosexuality is commonly recognized in the clinical field; the same is true in the case of female inversion and *active* female homosexuality. In such cases the choice of a member of the same sex for a love object appears to be simply an expression of the inverted sex-role identification of the individual, i.e., since *psychologically* the male invert perceives himself as a female or as predominantly feminine (and conversely with the female invert), he accordingly seeks the "opposite" sex as a love object. Largely on the basis of adult cases seen in clinical practice, it has been hypothesized that a primary determinant of sex-role inversion is excessive identification with the opposite-sex parent in earlier life, beginning in childhood and continuing thereafter.

---

[3] Acknowledgment is made to Dr. Robert R. Sears, Stanford University, for pointing out the importance of distinguishing between these concepts.

The backgrounds of adults with inverted sex roles (i.e., passive, feminine males, and active, masculine females) typically show an early and continuing preference for the behavior characteristic of the opposite sex and an aversion for that associated with their own sex. In the case of males, for example, a preference for wearing dresses, using cosmetics, wanting to be the mother in playing "grown-ups," and similar expressions of the feminine role, are traits that appear to be highly correlated with adult sexual inversion.

The relevance to the present study of the problem of passive male and active female homosexuality may be stated as follows: in a group of young children, do sex-role preference patterns exist that have been reported characteristic of the childhoods of male and female sexual inverts who are adult homosexuals?

THE PRESENT INVESTIGATION

The problem. The present study represents an attempt to investigate and analyze the nature and extent of young children's preference for objects and activities characteristic of their own or the opposite sex. The concept, *sex-role preference,* may be operationally defined in terms of the preferential responses of children to sex-typed objects and activities. In order to study this problem it was necessary to construct a scale that would facilitate the quantification of sex-role preference. The need for, and importance of, such a scale were specifically pointed out some 20 years ago by Terman and Miles, pioneers in the study of adult masculine-feminine role behavior, when they posed the following question: to what extent is the adult masculinity-femininity status of a person foreshadowed in childhood? These writers comment on this as follows: "The question is a very interesting one, but at present no answer can be given. *It cannot be answered until a more satisfactory M-F test has been devised for use with young children* . . . Because of the far-reaching effects which extreme M-F deviation may have upon an individual's social and sexual adjustment, the problem should be thoroughly investigated. . ." (24, pp. 15-16, italics added). The present study, in short, represents just such an attempt.

Subjects. Seventy-eight male and 68 female kindergarten children, ages 5-4 to 6-4 with a median age of 5-10 were used as subjects. These children were enrolled in a Denver public elementary school located in a middle-class section of the city. Analysis of the occupations of the children's fathers, based on a modification of the occupational rating indexes of Warner (27) and Centers (9), shows the present sample is made up largely of middle-class children, although within this class there are lower and higher levels.

Procedure. A sex-role preference scale, hereafter referred to as the *It Scale for Children* (ITSC),[4] was constructed and administered to the subjects. The previous work of Terman (23), Benjamin (6), and Rabban (19) was taken into consideration in devising the scale. As finally developed the scale consists of 36 picture cards, 3 by 4 inches, depicting various objects, figures, and activities commonly associated with masculine or feminine roles. The selection of items included in the scale was based on the contrasting behavior patterns socially identified with male or female roles. Thus, the kinds of objects and activities typical for boys in contrast to girls and vice versa, along with the more obvious differences between adult masculine and feminine roles to which the child is continually exposed—these considerations formed the basis for the content of the ITSC. In our culture, for example, boys normally play with trucks and trains, wear shirts and trousers, and grow up to use shaving articles; whereas girls typically play with dolls and dishes, wear dresses, and grow up to use cosmetics. Preferences for such male or female items are assumed to be indicative of preference for aspects of the masculine or feminine roles. Similarly, a boy who expresses a desire to be a girl, or a girl to be a boy, indicates a corresponding opposite sex-role preference. These examples illustrate what may be termed the operational validity of the ITSC. In other words, the logic and validity of the scale rests primarily on the assumption that what is socially regarded as, and actually associated with, masculine or feminine behavior is an adequate basis for defining sex-role patterns. The things and activities that one sex shows an interest in and liking for, in contrast to the other sex, would appear to be a valid criterion for defining sex-role behavior.

---

[4] Arrangements are being made to have this scale reproduced and made available to interested investigators for purposes of further research, standardization, and extension. The It Scale may be obtained from Psychological Test Specialists, Station E, Box 1, Louisville, Ky.

A child-figure drawing, referred to as "It," unstructured as to sex, was used by having each child make choices for It. The use of some kind of an animal figure was originally considered, but the It-figure seemed to offer a better means of facilitating the child's projections. The utilization of the It-figure represented an attempt to develop a more indirect, projective approach to the problem of measuring degrees of sex-role preference. Preliminary testing in which the child himself was asked directly to give his preference revealed that conformity to social expectations or pressure was more likely to be measured than the child's own preference. The utilization of the It-figure in the ITSC is an attempt to overcome this obstacle by focusing the entire testing situation on It rather than on the child directly. In using It the assumption is made that the child will project himself or herself into the It-figure on the basis of his or her own sex-role preference, and will attribute to It the child's own role preference. Thus, a girl who basically prefers the masculine role will tend to attribute this role to It, while the girl who prefers the feminine role will project such a preference to It, and the girl who is ambivalent or confused in sex-role preference will tend to give responses indicating a mixture of both masculine and feminine components.

In administering the scale, the child was seated in front of a small desk or table and to the left of the examiner, and was told simply that he or she was going to play a little game. The instructions to each child are reproduced below.

INSTRUCTIONS FOR ADMINISTERING THE IT SCALE FOR CHILDREN

Introduction: We are going to play a little game with this child here. See this child? Let's call this child "It." Let's play like the name of this child is It. O.K.? So this game will be about It. Here, you hold It. Now, we're going to show this child, whose name is It, some cards with pictures on them. [If child asks about the sex of It, say: "It's just a child, isn't it? Let's just say It's a child."]

1. *Toy pictures:* Look, here are some pictures of toys. These are all nice toys, aren't they? Yes. Now, let's play like It could play with all of these toys. Which toy would It like the best? Put It on the toy It likes the best . . . Now put It on another toy It likes. [Repeat until eight choices are made.] Are you having fun playing with It? Fine. Now we have some other pictures we're going to show It.

2. *Eight paired pictures (a)* Indians: Here are pictures of Indians. Let's play like It could be an Indian too. Put It on the Indian that It would rather be. *(b)* Clothes: Here are pictures of some clothes. Let's play like It could have any clothes It wanted. Put It on the clothes It would rather have. *(c)* Sewing-airplane: Here are pictures of

things to use in making an airplane and things to use in making a handkerchief. Which would It rather make, an airplane or a handkerchief? Put It on the one It would rather make. *(d)* Face articles: Here are pictures of things to use on our face. Now, let's play like It could play "grown-ups" and It had all these things to play "grown-ups" with. Put It on the one It would rather play "grown-ups" with. *(e)* Mechanical tools and household objects: Here are pictures of some objects to use in washing and ironing and some objects to use in fixing things that are broken. Let's play like It had all of these things to work with. Put It on the ones It would rather work with.

(f) Shoes: Here are pictures of some shoes. Now let's play like It could "dress up" and "play house." Put It on the shoes that It would rather "play house" with. *(g)* Children playing: Here are pictures of children playing together. Let's play like It could be in one of these pictures too. Put It on the picture It would rather be in. *(h)* Building tools and baking articles: Here are pictures of things to bake and cook with and things to build with. Let's play like It could work with all of these things. Put It on the things It would rather work with.

3. *Four child-figures:* Here are some pictures of children. Let's play like It could be any one of these It wanted to be. Put It on the one It would rather be.

A scoring sheet was used for each subject for recording purposes. Administration of the entire scale averaged about seven or eight minutes for each child.

The Toy Pictures Section of the ITSC included the following items:

Necklace
Tractor
Doll
Dump truck
Train engine
Purse
Gun (rifle)
High chair
Cradle
Racer
Dishes
Earthmover
Soldiers
Doll buggy
Knife (pocket)
Baby bath

The Eight Paired Pictures Section of the ITSC included the following picture pairs:

Indian princess—Indian chief
Trousers and shirt—Dress
Sewing materials—Airplane parts
Cosmetic articles—Shaving articles
Mechanical tools—Household objects
Men's shoes—Women's shoes
Girls playing—Boys playing
Building tools—Baking articles

The Four Child-Figures Section of the ITSC included the following:

Girl
Girlish boy (boy dressed as girl)
Boyish girl (girl dressed as boy)
Boy

The order of presenting the various sections and items of the scale, and their spatial arrangement, were randomly determined, with the restriction that male and female items be alternated.

For scoring purposes the following values were assigned to items comprising the scale:

Toy pictures: One point was given for each choice of a male toy picture and zero for each choice of a female toy picture. Each child made eight choices so that the range was from 0 (all female choices) to 8 (all male choices).

Eight paired pictures: Eight points were given for each male preference and zero for each female preference; since there are eight paired items, the range is from 0 (all female preferences) to 64 (all male preferences).

Four child-figures: Preference for the boy was given 12 points, for the girlish boy 8 points, for the boyish girl 4 points, and for the girl, zero.

The total range of the ITSC is from zero, an exclusively feminine score, to 84, an exclusively masculine score. A score of 42 would represent a relatively intermediate preference between masculine and feminine roles as defined in the present study. Deviations above and below a score of 42 would be in the direction of greater masculinity and greater femininity, respectively.

All results and conclusions of the present study are based on data obtained from use of the ITSC. The reliability of this scale was determined by the test-retest method with an interval of approximately one month between tests. Since the scatter diagrams of test-retest scores for boys and girls showed no evident departure from linearity, product-moment coefficients were computed. For boys the coefficient is .71, and for girls, .84. The fact that the distribution of scores for girls and particularly for boys does not meet the criterion of approximate normality necessitated the computation of rank-difference coefficients in order to test for significant departure from zero. These are .69 and .82 for boys and girls, respectively, both of which are significantly greater than zero at the .01 level of confidence. It is possible that sex-role preference at this age is somewhat labile; if so, this may have a lowering influence on the obtained reliability coefficients. The lower reliability in the case of boys compared to girls would appear to be due at least in part to the relatively greater homogeneity of the former. On the whole, however, the reliability of the ITSC seems reasonably adequate and compares favorably with other evaluative instruments that have been used in measuring various psychological characteristics of young children.

Results and Conclusions

Group sex-role patterns. An assumption underlying the present study is that sex-role preference patterns exist in young children—that young boys as a group identify with models, activities, and goals that are socially defined as masculine, while young girls as a group identify with those that are socially defined as feminine. Thus, very different patterns in boys compared to girls might be expected. As may be noted in Fig. 1 and Table 1, large and significant differences occur between boys and girls, indicating the existence of definite, relatively dichotomous sex-role preference patterns in boys and girls. The mean difference of 28 points and the median difference of 33 points between boys and girls are highly significant. The upper quartile for girls is approximately the same as the lower quartile for boys. Although the two distributions overlap, as would be expected unless boys and girls were completely dichotomous with respect to sex-role preference, the differences between them are clearly evident. Further suggested is the fact that these children show considerable variability in their preferences, girls being significantly more variable than boys.

Mixed or ambiguous sex-role patterns. A number of children in both groups show a mixed or confused preference pattern, indicating acceptance of components of both the male and female roles. This tendency is about twice as frequent in girls as in boys, in that 34 per cent of the girls and only 18 per cent of the boys score at or close to the intermediate score of 42. In addition, a sizable number of both groups are within the score range indicating considerable mixed preference. These results may have interesting implications with respect to the hypothesis of Mowrer (16) and others that early and *continuing* confusion in sex-role learning is a basic source of adult maladjustment. Further research, ideally of a

TABLE 1

GROUP SCORES, VARIABILITY, AND DIFFERENCES ON THE IT SCALE FOR CHILDREN
(Total Sample, Boys and Girls)

	N	Q1	Q3	Median	Mean	CR	*SD*	*F*
Boys	78	56.29	80.75	71.00	66.36	8.42*	16.67	1.79**
Girls	68	19.50	55.50	37.83	38.40		22.28	

* Significant beyond the .001 level.
** Significant beyond the .01 level.

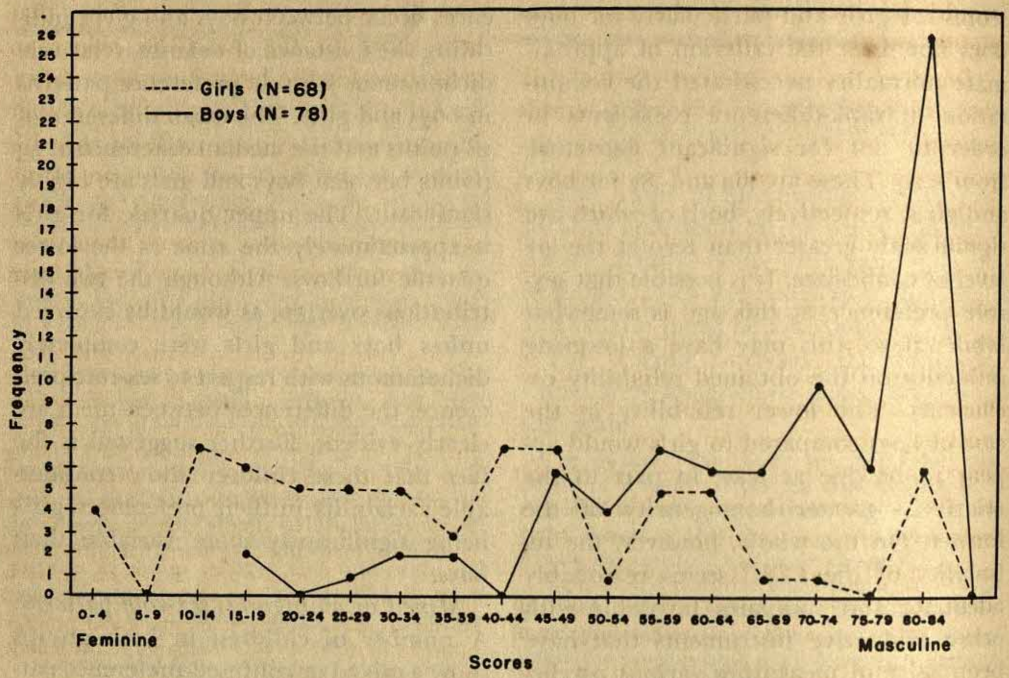


FIG. 1. Distribution of scores on the It Scale for Children — Total sample (boys and girls).

longitudinal nature, is very much needed in this area. The crucial question here would pertain to the constancy of sex-role patterns, i.e., whether these children who show ambivalence at six years of age develop as such into adulthood. These results also suggest that to the extent sex-role preference patterns involving a mixture of both masculine and feminine components are indicative of uncertainty and ambivalence, more girls than boys experience conflicts in their sex-role development. Another source of conflict, in addition to the mixture of both roles, lies in the extent to which there is a discrepancy between the child's role identification and his or her role preference. The question, however, as to how much the ITSC taps role *preferences* in relation to how much it taps role *identifications* is one that does not appear answerable from the present data. Related to the problem of greater sex-role ambiguity in girls is an observation made by Seward,[5] to the effect that such apparent ambiguity of girls may be a function of their "minority" status. "In all minority groups there is necessarily a double identification—with the dominant group as well as with the specific minority. In the case of females you would expect some identification with the male role representing the dominant group as well as with the female 'minority.' " A further point made by Seward in this same connection is that, contrary to the hypothesis of Mowrer that sex-role ambiguity is related to maladjustment, such ambiguity may actually involve a flexibility of sex-role in girls, since the feminine role is in such a state of flux in our culture. "In other words, it might be better for a child to have some ambiguity and mixed role identification in order to avoid the conflict which would be inevitable if the role were rigidly set in early life."

---

[5] Personal communication from Dr. Georgene H. Seward, University of Southern California.

Inverted sex-role patterns. Some children in both groups show considerable inversion of sex roles, i.e., a strong opposite sex-role preference. This tendency, which is evident in Fig. 1, is more frequent and more pronounced in girls than in boys. Psychoanalytic and other personality theorists hold that such reversals or inversions of role constitute a potential source of adult, male passive and female active homosexuality. In this connection there would seem to be important implications here with respect to the continuation of childhood sex-role inversion in relation to later adult adjustment. Should future research confirm the tendency observed by Terman and Miles (24) in preadolescents and young adults, namely, a relative constancy of masculinity-femininity patterns over a period of years, results such as those obtained here may come to have considerable predictive value in this area, and eventually be of help in suggesting directions for re-educative procedures and preventive mental hygiene.

Masculine role preference in boys compared to feminine role preference in girls. A third outcome of sex-role identification, postulated by Mowrer (16), is the normal adult who in childhood desired and accepted the role socially appropriate to his or her sex. Figure 1 suggests that at this age boys fare much better in this respect than do girls, at least in terms of sex-role preference. Although some 20 out of 68 girls show a marked preference for femininity, the others are in varying degrees removed from clear-cut feminine preference. On the other hand, a clear majority of boys reveal a decided preference for the masculine role. Thus one of the most striking findings in the present study is the comparatively greater preference that boys show for the masculine role than girls show for the feminine role, a difference that is large and significant. Another indication of the magnitude of this difference is the fact that the mean of deviations of girls (38.40) from an exclusively feminine score is more than twice the mean deviation of boys (17.64) from an exclusively masculine score.

These results are in agreement with those reported by Rabban (19, p. 125), who found that "Middle class girls significantly lag in sex-appropriate choices from the age of five years." Hurlock (14, p. 121) interprets Rabban's finding here as indicating boys are "more clearly aware" of appropriate sex-role behavior than are girls. However, it may be simply that boys prefer aspects of the masculine role more than girls prefer aspects of the feminine role. Furthermore, such findings are not out of keeping with the widespread assumption of greater status and prestige attached to the male role compared to the female role in our culture.

It is also obvious that these results are consistent with both the Freudian theory of penis envy and the Adlerian theory of masculine protest, although it is not possible in the present study to evaluate the comparative validity of Freudian versus Adlerian constructs in accounting for these results; and, of course, the present data do not necessarily substantiate the underlying logic or validity of either position. Furthermore, despite the consistency of present findings with the hypothesized dissatisfaction of girls with the feminine role, the question remains as to those girls who do obtain predominantly feminine scores. Are they as proud and satisfied with their femininity as the majority of boys ostensibly are with their masculinity? Or is it that they have simply resigned themselves to the "fate of being feminine" at an earlier age than most girls in this age group? The entire problem concerning the dynamics of sex-role development is one that needs exploration and experimental study.

In a sense the present results appear to run counter to the formulations of Parsons (17, 18) and Gorer (12) to the effect that the little girl earlier and more readily acquires her sex role than is true for boys. In view of the distinction between identification and preference, however, another interpretation is possible, namely, that the girl subjects may very well be fully aware of and identified with the feminine role but nevertheless *prefer* varying aspects associated with the masculine role. Thus, it may be that the sex-role identifications of girls are *not* "more confused and less clear-cut" than those of boys (7, p. 117), but rather that girls consciously *prefer* aspects of the masculine role even though their unconscious identifications are basically feminine. In addition, the present data might be interpreted as bearing out the view of Gorer (12) and Parsons (18) that boys are *compulsively* masculine, i.e., will show exaggerated tendencies in this direction, in part because they are *less* sure in their sex-role identification than girls and in part because they sense the less favored position of the female. Thus boys might be expected to be overtly assertive and impelled toward overcompensation in their masculine strivings.[6] And for the same reasons, girls, being more secure in their feminine identification but at the same time being aware of greater male prerogatives, might be expected to express more freely preferences for certain aspects of the masculine role. These considerations seem to lead to the assumption that in sex-role development, girls compared to boys would have an easier course in terms of role *identification,* but at the same time experience more difficulty relative to role *preference;* and, conversely, boys compared to girls would have a more difficult course relative to identification, but would experience an easier one in terms of role preference. Thus, since the present results were obtained in a study of sex-role *preference,* this fact would appear consistent with the foregoing analysis.

A somewhat different but nevertheless relevant comparison may be made here with the findings of Simpson (22) who was concerned with parent preferences of young children. She found pronounced mother preference in both boys and girls at each age from five to nine with the exception of *five-to-six-year-old girls,* 60 per cent of whom preferred the father. "The most striking fact . . . is that the girls of the five-to-six-year-old group do not conform to the trend of all the other age groups" (p. 24). "It is hard to explain this exception satisfactorily" (p. 67). The fact that the age of this group of girls and that of the present group coincide has interesting implications for future research. The problem of age is immediately evident—do girls *after* about the sixth year show greater preference for the female role in line with their greater preference for the mother? Is the strong trend of masculine preference noted in the present study related to the finding that girls at this age tend to prefer the father rather than the mother? Another problem raised here is the fact that contrary to girls of this age, boys much more frequently prefer the mother but at the same time strongly prefer the masculine role. One factor that will have to be better clarified in future study in this area is that of "parent preference." What is meant when a child is asked, as in Simpson's study, "Which parent do you like best?" It seems reasonable to suppose that in some ways and for some purposes, the child may prefer or like one parent better; in other ways and for other purposes the other parent may be preferred. Thus, for example, the boy may prefer the mother as the more loving parent or in the sense of a love object, but at the same time prefer the father as the more esteemed parent or in the sense of an identification model. The need is to differentiate the term "parent preference" as to identification model, love-object choice, rewarding agent, punishing agent, and in other ways significantly related to the role of the parents in the life of the child.

Item analysis of the It Scale for Children. An analysis of the items and sections of the ITSC provides further information relative to the differences between boys and girls in their sex-role preference patterns.

Toy pictures section. In the development of the ITSC it was assumed that toy objects commonly associated with boys and those commonly associated with girls constitute one source of difference in sex-role preference. To the extent that boys prefer the masculine and girls the feminine role, differences in their choices of play objects connected with such roles should be evident. Results listed in Table 2 indicate that on the whole this prediction is confirmed. A greater percentage of girls than boys prefer each of eight female toy objects, while a greater percentage of boys than girls prefer each of eight male toy objects.

Despite the expected consistent sex differences noted above, there are interesting differences in the relative extent of boys' choices of male toys compared to girls' choices of female toys. For example, since each child made eight choices, it might be expected that a rank order of

---

[6] Zilboorg (28, p. 281) suggests that our patriarchal culture pattern, consisting of male dominance and overly valued, assertive masculinity, is in part based on a defensiveness and reaction on the part of males to the powerful, pervasive influence of the maternal figure in their formative years of personality development.

TABLE 2

RANK ORDERS, PERCENTAGES, AND DIFFERENCES OF CHOICES OF BOYS AND GIRLS ON THE TOY PICTURES SECTION OF THE IT SCALE FOR CHILDREN

Toy item	Boys' rank	Girls' rank	Boys' %	Girls' %	Chi square	Level of confidence
Necklace	13	12.5	28	40	2.17	Not significant
Tractor	3	14	77	38	22.45	Beyond .001
Doll	11	1.5	35	71	18.82	Beyond .001
Dump truck	4	9	71	50	6.42	Beyond .02
Train engine	7	11	59	41	4.61	Beyond .05
Purse	13	15	28	37	1.22	Not significant
Gun (rifle)	1	4	87	63	11.43	Beyond .001
High chair	9	5	46	60	2.92	Not significant
Cradle	16	16	12	19	1.62	Not significant
Racer	5	7.5	68	51	4.13	Beyond .05
Dishes	10	1.5	40	71	13.85	Beyond .001
Earthmover	2	3	81	66	4.02	Beyond .05
Soldiers	8	10	54	49	.41	Not significant
Doll buggy	15	12.5	22	40	5.54	Beyond .02
Knife (pocket)	6	6	65	53	2.33	Not significant
Baby bath	13	7.5	28	51	8.25	Beyond .01
Total choices of male toys			70	51		
Total choices of female toys			30	49		

the 16 toys would show that ranks 1 to 8 would include only male toys in the case of boys and only female toys in the case of girls. This expectation is confirmed for boys but not for girls. For the latter, four out of the first eight ranks are female (doll, dishes, high chair, and baby bath) but four are male (earthmover, gun, knife, and racer) toy items. This result is consistent with the over-all finding that girls do not show preference for components of the feminine role to the extent that boys show preference for the masculine role. Another example of this is the considerable and highly significant difference in the total number of male and female toy choices of boys and girls. Thus 70 per cent of all choices of boys are masculine, whereas only 49 per cent of the choices of girls are feminine. It is recognized that these results may have been influenced in the sense that the male items, *apart* from their sex-role connotation, may have been more desirable than the female items. Parallel to this is the possibility that the drawings themselves were such as to enhance the attractiveness of boy toy items over the girl items. Although it is not possible to determine the extent, if any, that these factors may have been operative in the present study, careful attempts were made throughout the construction of the scale to negate their influence.

These results may be compared to those obtained by Rabban (19) in his study involving the use of actual toy *objects* in which he asked each child *directly* which toy he or she liked. Rabban found large and significant differences between boys and girls in the case of each of 16 toy objects. Although otherwise consistent with Rabban's findings, this is in contrast to the present finding of 6 out of 16 toy picture cards (necklace, purse, high chair, cradle, soldiers, and knife) that did not significantly differentiate between boys and girls. The gun and earthmover are the two most popular choices of boys, while the doll and dishes are most popular with girls. This is in line with the findings of Terman and Miles (24, p. 12) who found "shooting" and "work with machinery" among the most masculine and "play dolls" and "cook and play house" among the most feminine activities of children.

The failure in the present instance to obtain significant differences on every item may be due to one or more of several factors: (*a*) geographical or socioeconomic differences between samples in the two studies, (*b*) the use of actual toy objects rather than pictures of toys, (*c*) the tendency of some children in the present study to

Bureau Ednl. Psy. Research
AINING COLLEGE
Dated ....

make two or three premeditated choices and thereafter to choose the remaining items indiscriminately or without any apparent forethought, and *(d)* asking the child directly his or her preference in contrast to a more indirect approach that utilizes an unstructured child figure. The first factor may be influential here inasmuch as Rabban's total sample included a considerable number of lower social class subjects: thus, the two samples differ somewhat in their socioeconomic composition. That geographical or regional variations produce differences is also possible, although it is quite likely that the toy items used are common to practically every section of the country. The second factor, toy objects versus pictures, may produce different choices although it seems reasonable to assume a high correlation between them. It is not possible to determine to what extent the third factor influenced the results, although a slight change in the manner of administering the toy item section would probably eliminate this as a source of failure to obtain significant differences between boys and girls on all items. For example, instead of presenting all 16 cards at one time and having the child make eight consecutive choices, the items could be presented in groups of four or eight. The influence of the fourth factor should be tested not only because of the differences between Rabban's results and the present data, but also in view of the implications for perceptual theory and methodological problems in the field of child and personality study. Specifically, results obtained in the present study involving the It figure should be compared with results based on direct focus on the child himself. If the indirect approach used here facilitates a more unrestricted, uninhibited expression of sex-role preference in the child, differences in the two approaches with suitable controls should reflect this fact.

Eight paired pictures section. Although toy item choices in themselves constitute one basis for defining sex-role preference, it was felt that a more adequate appraisal could be obtained if paired items were included, some of which were directly connected with masculinity-femininity in the adult sex roles as well as childhood sex roles. For this reason the Eight Paired Pictures section was included in the ITSC. The results in Table 3 show large and significant differences between boys and girls on each of these paired items. The number of boys choosing the masculine alternative varies from 71 per cent on the boys-girls playing item to 91 per cent on the cosmetic-shaving articles item, whereas the number of girls choosing the feminine alternative varies from 40 per cent on the sewing-airplane item to 69 per cent on the building-baking articles item. These results point once again to the fact that girls do not show the same preference for aspects associated with their sex role as is true for boys. In fact, on five out of the eight items, half or more of the girls express a preference for the masculine alternative. For example, when asked which Indian It would rather be, 51 per cent of the girls indicated the male Indian. When asked which It would rather make, an airplane or a handkerchief, 60 per cent expressed preference for the former. And when asked which things, cosmetics or shaving articles, It would rather play "grown-ups" with, 50 per cent chose the latter. These percentages stand in sharp contrast to those obtained in the case of boys on these same items, of whom 86, 88, and 91 per cent, respectively, indicated a preference for the masculine alternative.

The three items on which girls show the largest preference for the feminine alternative are: cooking-baking articles, womens' shoes, and girls' dress. Even here, however, it is evident that a sizable minority of the girls (31, 34, and 38 per cent) make masculine choices.

In the case of boys it is interesting that the smallest number of feminine responses occur to items containing cosmetics and sewing articles. Certainly one of the least common patterns of males in our culture is the use of lipstick, rouge, needle, thread, etc. On the other hand, equally as uncommon is the use of razor and shaving brush among females; yet, half of the girls indicate a preference for these articles over cosmetics.

TABLE 3

PERCENTAGES AND DIFFERENCES OF CHOICES OF BOYS AND GIRLS ON THE EIGHT PAIRED PICTURES SECTION OF THE IT SCALE FOR CHILDREN

	Boys		Girls		Chi square	Level of confidence
	Male choice %	Female choice %	Male choice %	Female choice %		
Indian princess Indian chief	86	14	51	49	20.46	Beyond .001
Trousers and shirt Dress	77	23	38	62	22.45	Beyond .001
Sewing materials Airplane parts	88	12	60	40	15.50	Beyond .001
Cosmetic articles Shaving articles	91	9	50	50	30.26	Beyond .001
Mechanical tools Household objects	82	18	49	51	18.13	Beyond .001
Men's shoes Women's shoes	76	24	34	66	25.80	Beyond .001
Girls playing Boys playing	71	29	50	50	6.42	Beyond .02
Building tools Baking articles	74	26	31	69	27.65	Beyond .001
Totals: eight paired items	81	19	45	55		

The item involving girls playing together and boys playing together is of interest in itself, in view of several studies concerned with the sex of playmate choices in children. In this connection Campbell (8) has concluded that children in the 5-to-8-year age group show no definite sex preference; however, clear-cut preferences for playmates of the same sex, even in younger children, have been reported (14). The present results in the case of boys suggest that they more frequently prefer to play with boys than girls. Girls, however, are evenly divided, half expressing preference for male and half for female playmates. This tendency in girls is in line with the overall finding that more girls than boys tend to prefer the opposite sex role, one part of which seems to involve a preference for the opposite rather than the same sex playmates. In fact, one of the reasons for including this item in the scale is the finding that consistent preference for opposite sex playmates is correlated with an opposite sex-role preference (24).

Finally, the magnitude of difference between boys and girls with reference to degree of masculinity-femininity is seen in the fact that 81 per cent of the total choices of boys are for the masculine alternative of the Eight Paired Pictures section, while only 55 per cent of the choices of girls are for the feminine alternative.

Four child-figures section. The last section of the ITSC, the Four Child-Figures section, involves the use of four child figures to which the child responds by indicating the one that It would rather be. This item gets at the problem of sex-role preference most directly in the sense the child's response here suggests not only his or her desired role but the preferred sex itself. In short, the child does not simply indicate a preference for objects or activities associated with one sex or the other; he or she expresses, via It, a preference for being a male or female. The results once again suggest that girls do not show nearly the preference for being female as boys show for being male. Data contained in Table 4 indicate that more than twice as many girls as boys express a preference for being a person of the opposite sex. Whereas 31 per cent of the girls indicate that if It could be any kind of a child, It would rather be a boy, only 12 per cent of the boys express a preference for It wanting to be a girl. Responses to the mixed figures, girlish-boy (boy dressed as girl) and boyish-girl (girl dressed as boy) show that about one-fourth of the girls choose the masculinized girl whereas only one-tenth of the boys choose the feminized boy. Although this is consistent with the over-all tendency of girls to respond more in the direction of masculinity than of boys toward femininity, the fact that it is culturally permissible for girls to dress as boys but not for boys to dress as girls here would appear to be a relevant factor. In a sense the same may be said for other aspects of the male and female roles, i.e., there is less social disapproval involved in girls participating in masculine activities than in boys following feminine pursuits. Nevertheless there is a large and unmistakable sex difference on this section of the scale that suggests being male is favored over being female.

TABLE 4

PERCENTAGES AND DIFFERENCES OF CHOICES OF BOYS AND GIRLS ON THE FOUR CHILD-FIGURES SECTION OF THE IT SCALE FOR CHILDREN

Child figure	Boys' %	Girls' %	Chi square	Level of confidence
Girl	12	38	23.20*	Beyond .001
Boyish Girl	10	26		
Girlish Boy	9	4		
Boy	69	31		

* Because of small frequencies in several cells, chi square computed by combining Girlish Boy and Boyish Girl Choices for both groups.

This rather pronounced trend is consistent with the findings of Terman (25, p. 259) and Landis, Landis, and Bolles (15, p. 78) who posed the question: "Have you at some time in your life wished that you were of the opposite sex?" Terman found in a sample of 792 couples that about 31 per cent of the women but only about 2½ per cent of the men indicated they had wished they were of the opposite sex; and among 549 intellectually gifted women or wives of gifted men and 547 intellectually gifted men or husbands of gifted women, Terman[7] found that 42 per cent of the women but only about 5½ per cent of the men indicated they had sometimes wished they were of the opposite sex. And in the study by Landis, Landis, and Bolles of 295 women, 61 per cent reported having wanted to be a boy at some time in their lives.

Thus, in our culture, it seems safe to conclude there are many more girls who prefer to be male than there are boys who prefer to be female.[8] However, the distinction between role preference and role identification should be kept in mind here. The fact that a large percentage of females express a *preference* for being male does not mean that the basic sex-role *identification* of such girls and women is masculine. If this were the case much more sexual inversion among women would exist than is actually found.

---

[7] From data on gifted subjects in the personal files of Dr. Lewis M. Terman, Stanford University.

[8] Dr. Ernst G. Beier, University of Utah, in a personal communication with the writer and based on a study of the parental identifications of college students (5), suggests that our entire culture pattern is going in the direction of a matriarchial society in the sense that girls and women seem rapidly to be adopting masculine ways, goals, and behavior, and are assuming an increasingly dominant role. Such examples as the trend toward slim bodies, boyish figures, masculine hairdo's, the wearing of shirts and pants, as well as cultural and professional pursuits commonly identified with the male role, are cited in this connection. The findings of the present investigation are entirely consistent with these observations by Beier, and suggest that such masculine strivings in girls have an early beginning.

It might be suggested in this connection, that sex-role inversion is a function *both* of basic identification with and a basic preference for the role of the opposite sex. Thus, for example, a female sexual invert is one who has made a predominantly masculine identification as well as one who strongly prefers the masculine role, and vice versa for the male invert. This is the pattern found in cases of active female homosexuality and, conversely, in cases of passive male homosexuality.[9] In any event, results of the present study suggest that sex-role inversion patterns exist in some children in early childhood. The possibility of a continuation of such tendencies into adulthood is a problem that merits careful consideration and further exploration. The implications are obvious in view of the widespread incidence of homosexuality in our culture.

Test-retest differences in sex-role patterns. For purposes of determining reliability, the ITSC was readministered about one month after original testing. Results show essentially the same differences and patterns obtained in original testing and discussed above. There is, however, a slight but insignificant mean shift of boys (1½ points upward) and of girls (3 points downward), reflecting, respectively, a slightly stronger masculine preference of boys and feminine preference of girls.

Reaction time in sex-role choices. The reaction time of subjects in making choices to sections of the ITSC was recorded, on the assumption that prolonged responses would indicate ambivalence or confusion. No significant sex differences in reaction time, however, were found in responses to the combined Eight Paired Pictures and in response to the Four Child-Figures sections. Thus, if girls do experience greater equivocality than boys in making component role choices, reaction time in the present instance fails as an indication that such is the case.

Sexual identity attributed to the It-figure. Responses to the ITSC were based on the use of an ambiguous-sexed drawing of a child named It. Since the objective was to use a figure that has neither male nor female characteristics as such, the child's perception of It as a boy or as a girl would appear to be governed primarily not by objective stimulus dimensions but by the child's own aspirations and desires. The assumption is made that children tend to perceive the It-figure in terms of their own self-image or preferred self-image and thereby attribute to It sex-role components that are actually operative in themselves.

At the conclusion of the second testing session each child was asked to think of a name for It ["Now imagine a real name for It, just think of some name It would like to have"]. As may be noted in Table

---

[9] A distinction should be made between sex-role inversion, i.e., the adoption of the overt and covert behavior of the opposite sex, and homosexuality, i.e., sexual activity between two members of the same biological sex. The former appears to involve an inversion of the entire personality in the sense that the invert is one whose thoughts, perceptions, attitudes, fantasies, feelings, interests, and actions are typical of the opposite sex. Homosexuality, then, would be simply another manifestation of the individual's inversion. However, homosexuality has correlates, forms, and manifestations other than those involved in role inversion. In short, while sex-role inversion typically involves homosexuality, not all homosexuality is based upon sex-role inversion.

TABLE 5

DIFFERENCES BETWEEN BOYS AND GIRLS IN THE SEX OF NAME ATTRIBUTED TO THE IT FIGURE

	Boys (*N* =77)	Girls (*N* =66)	Chi square	Level of confidence
Male name attributed to It	66	36	16.90	Beyond .001
Female name attributed to It	11	30		

5, 85 per cent of the boys gave It a male name, while only 45 per cent of the girls gave It a female name. Stated differently, only 15 per cent of the boys assign a female name while 55 per cent of the girls assign a male name to It. Again, there is the suggestion here of greater preference for the male compared to the female role. In a general sense this trend is consistent with the finding of Amen (2), who found in a study of preschool children that, in assigning sexual identity to a sexless black and white silhouette picture of a child, the "same-sex choices were less clear for girls than for boys, since many of the girls referred to the child merely as a 'baby'" (p. 357). A point that should be mentioned with reference to the present finding is that concerning the It-figure drawing. Despite the specific attempt to neutralize sex traits in the drawing of It, there is the possibility that the figure may be more readily perceived as male than female, although there was no indication that such was the case in testing the present group of children. In any event this possible source of bias should be investigated in future research.

An interesting aspect arises in connection with the differences in ITSC scores between boys who perceive It as male or female and between girls who perceive It as male or female. A reasonable prediction would be that boys naming It a male will score more masculine than boys naming It a female; and, likewise, girls who name It a female will score more feminine than girls who name It a male. Actually, as may be noted in Tables 6 and 7 there is a significant 13-point mean difference in boys and a 27-point mean difference in girls, which indicate that boys and girls who gave It a name consistent compared to those who gave It a name inconsistent with their own sex are, respectively, more masculine and more feminine in role preference.

Social class in relation to sex roles. Although the present sample constituted a relatively homogeneous middle-class group, it was subdivided on the basis of the father's occupational status into three smaller groups that might be termed: upper-middle, middle, and lower-middle class. Comparisons were made among these three classes; no significant mean differences were found. It would appear that the relative homogeneity of children in the present study with respect to overall middle-class status is related to the uniformity of sex-role preference scores when intraclass divisions are made. The findings of Rabban (19), however, suggest that definite differences in sex-role behavior do exist between children from *different* social classes. This problem of sex-role preference and identification of children of widely different social class statuses is an important area for future research.

Siblings in relation to sex roles. A final comparison was made between the ITSC scores of children who had, or did not have, siblings of the same and/or opposite sex. This analysis was made on the assumption that the presence or absence of siblings might have a bearing on the sex-role preference of these children.

TABLE 6

DIFFERENCES BETWEEN MEAN SCORES ON THE IT SCALE FOR CHILDREN OF BOYS WHO ATTRIBUTE A MALE OR A FEMALE NAME TO THE IT FIGURE

	N	Mean	*SD*	*CR*	Level of confidence
Boys who give It a male name	66	69.76	14.72	2.34	.02
Boys who give It a female name	11	57.09	16.10		

The results do not support this assumption in the case of girl groups, in which no significant differences were observed. In the case of boys, however, a 7-point mean difference, significant at the .05 level, is obtained between boys who had only sister(s) and those who had only brother(s), the former group being less masculine in their responses than the latter group. This trend appears reasonable in the sense that a boy who has one or more sisters normally would be exposed to a greater degree of feminine patterns than a boy who has one or more brothers. A difference of about the same magnitude and in the same direction was also found between boys having only sister(s) and those having both brother(s) and sister(s). This difference approaches significance (.10 level). These results suggest the possibility of some degree of feminization in the case of boys who have only sister(s) as siblings. Such results must be interpreted with caution, however, since out of 12 possible comparisons, only these two differences reach or approach significant levels. In addition, there is no corresponding trend evident in the case of girls. A factor not considered in these comparisons is whether or not the child's sibling(s) were older or younger than he or she. It would seem reasonable, for example, that the influence of an older sister on a boy's developing sex-role pattern would be different from the influence of a younger sister. In any event, the position of a child in the family with respect to the presence or absence of siblings of the same or opposite sex undoubtedly has a formative influence on his or her sex-role development and should receive further study. No less important in this connection would be the degree to which the parents, through their attitudes and feelings, approve and encourage the child in his or her sex-role development and behavior. In short, the relationship between the structural and dynamic constellation of the family and the sex-role preference of children is a crucial problem that merits further and more exact analysis.

SUMMARY

The present study was concerned with

TABLE 7

DIFFERENCES BETWEEN MEAN SCORES ON THE IT SCALE FOR CHILDREN OF GIRLS WHO ATTRIBUTE A MALE OR A FEMALE NAME TO THE IT FIGURE

	N	Mean	*SD*	*CR*	Level of confidence
Girls who give It a male name	36	46.64	22.25	5.86	Beyond .001
Girls who give It a female name	30	19.40	14.79		

the nature, patterns, and differences of sex-role preference in young children. Seventy-eight male and 68 female middle-class kindergarten children, ages 5-4 to 6-4, were used as subjects. A sex-role preference scale, the *It Scale for Children* (ITSC), consisting of 36 picture cards, 3 by 4 inches, depicting various objects, figures, and activities associated with masculine or feminine roles, was constructed and administered. A child-figure drawing named It, unstructured as to sex, was used, and each subject was asked to make choices for It. All conclusions of the present study are based on responses to this scale, the test-retest reliability of which (at an interval of one month) is .69 for boys and .82 for girls.

The findings are as follows: (*a*) large and significant differences occur between boys and girls, suggesting the existence of definite, relatively dichotomous sex-role patterns in young children. (*b*) A number of children in both groups showed a mixed preference pattern, indicating acceptance of components of both the male and female roles; this tendency was about twice as frequent in girls as in boys. (*c*) Some children in both groups showed a strong opposite-sex role preference; this tendency is more frequent and more pronounced in girls than in boys. (*d*) Boys show significantly greater preference for the masculine role than girls show for the feminine role. (*e*) Evidence for the assumption of greater prestige and value in the male compared to the female role in young children was found in the analysis of separate items of the ITSC. Among 16 sex-typed toy object items, 70 per cent of all choices of boys are masculine; only 49 per cent of the choices of girls are feminine. On paired items reflecting aspects of masculinity-femininity, from 71 to 91 per cent of the boys chose the masculine alternative, whereas from 40 to 69 per cent of the girls chose the feminine alternative. On the section that involved a choice between wanting to be a boy or a girl or a mixture of both, only about one boy in 10 projected a preference for the female, while nearly one girl in three expressed a preference for the male. (*f*) Retesting a month following original testing showed essentially the same differences and patterns obtained in the first testing. (*g*) No significant sex differences in reaction time in making sex-role choices were found (*h*) Eighty-five per cent of the boys, but only 45 per cent of the girls, gave the It figure a name consistent with their own sex. (*i*) No significant differences were found between upper-middle, middle, and lower-middle class subgroupings of the total group. (*j*) Boys who have only sister(s) scored somewhat more feminine than boys having only brother(s); and there was a similar tendency among the boys having both brother(s) and sister(s). No significant differences were found relative to this variable among girl groups.

Hypotheses and implications of the present investigation were briefly reviewed in terms of the nature of identification, sex differences in the acceptance of sex roles, culturally determined conflicts in sex-role behavior, and homosexuality in relation to sex-role development.

REFERENCES

1. Adler, A. *Understanding human nature.* Garden City: Garden City, 1927.
2. Amen, Elizabeth W. Individual differences in apperceptive reaction: a study of the response of preschool children to pictures. *Genet. Psychol. Monogr.,* 1941, **23,** 319-385.
3. Anastasi, Anne, & Foley, J. P., Jr. *Differential psychology.* (Rev. Ed.) New York: Macmillan, 1949.
4. Bach, G. R. Father-fantasies and father-typing in father-separated children. *Child Develpm.,* 1946, **17,** 63-80.
5. Beier, E. G., & Ratzeburg, F. The parental identification of male and female college students. *J. abnorm. soc. Psychol.,* 1953, **48,** 569-572.
6. Benjamin, H. Age and sex differences in the toy-preferences of young children. *J. genet. Psychol.,* 1932, **41,** 417-429.
7. Blum, G. S. *Psychoanalytic theories of personality.* New York: McGraw-Hill, 1953.
8. Campbell, Elise H. The social-sex development of children. *Genet. Psychol. Monogr.,* 1939, **21,** 461-552.
9. Centers, R. *The psychology of social classes.* Princeton: Princeton Univer. Press, 1949.
10. Ferguson, L. W. The cultural genesis of masculinity-femininity. *Psychol. Bull.,* 1941, **38,** 584-585. (Abstract)
11. Freud, S. *New introductory lectures on psychoanalysis.* London: Hogarth, 1946.
12. Gorer, G. *The American people.* New York: Norton, 1948.
13. Heston, J. C. A comparison of four masculinity-femininity scales. *Educ. psychol. Measmt,* 1948, **8,** 375-387.
14. Hurlock, Elizabeth B. *Developmental psychology.* New York: McGraw-Hill, 1953.
15. Landis, C., Landis, Agnes T., & Bolles, M. Margaret. *Sex in development.* New York: Hoeber, 1940.
16. Mowrer, O. H. *Learning theory and personality dynamics.* New York: Ronald, 1950.
17. Parsons, T. Age and sex in the social structure of the United States. *Amer. sociol. Rev.,* 1942, **7,** 604-616.
18. Parsons, T. Certain primary sources and patterns of aggression in the social structure of the western world. *Psychiatry,* 1947, **10,** 167-181.
19. Rabban, M. Sex-role identification in young children in two diverse social groups. *Genet. Psychol. Monogr.,* 1950, **42,** 81-158.
20. Seward, Georgene H. Cultural conflict and the feminine role: an experimental study. *J. soc. Psychol.,* 1945, **22,** 177-194.
21. Seward, Georgene H. *Sex and the social order.* New York: McGraw-Hill, 1946.
22. Simpson, Margarete. Parent preferences of young children. *Contrib. Educ.,* 652, Teach. Coll., Columbia Univer., 1935.
23. Terman, L. M., *et al. Genetic studies of genius.* Vol. 1. Mental and physical traits of a thousand gifted children. Stanford: Stanford Univer. Press, 1925.
24. Terman, L. M., & Miles, Catharine C. *Sex and personality.* New York: McGraw-Hill, 1936.
25. Terman, L. M. *Psychological factors in marital happiness.* New York: McGraw-Hill, 1938.
26. Tyler, Leona E. *The psychology of human differences.* New York: Appleton-Century-Crofts, 1947.
27. Warner, W. L., Meeker, Marchia, & Eells, K. *Social class in America.* Chicago: Science Research Associates, 1949.
28. Zilboorg, G. Masculine and feminine. *Psychiatry,* 1944, **7,** 257-296.

(Accepted for publication November 11, 1955)

Vol. 70, No. 15 Whole No. 422, 1956

Psychological Monographs: General and Applied

Conditioning as Emotional Sensitization and Differentiation[1]

ROSCOE A. DYKMAN[2]
School of Medicine, University of Arkansas
W. HORSLEY GANTT and JOHN C. WHITEHORN[3]
Johns Hopkins Hospital

INTRODUCTION

STUDENTS of the learning process have been much impressed by the probable significance of an emotional factor which might be variously called emotional excitation, anxiety, vigilance, or alertness (8, 11, 13, 18). Work on cardiac reactivity in this laboratory over the past 17 years has indicated that the heart is an excellent organ for the study of emotional excitation.

The purpose of the present research was to study certain cardiovascular aspects of motor learning in a conditioning situation where the motor conditional reflex (cr) could be contrasted with the associated cardiac changes at various stages of the development of both responses.

The operational aim of the research was the study of the latency and the amplitude of the cardiac and motor responses to stimuli signalizing different intensities of electrical stimulation. This is an extension of the work originally done in this laboratory relating the quantity of food reinforcement to the size of conditional and unconditional salivary reflexes (4).

PROCEDURE

Pavlovian methodology was employed. Four dogs, one to two years old, two male and two female, served as subjects. The experimental room consisted of a sound-shielded enclosure (camera) that separated the animal from adventitious stimulation and the operation of the recording apparatus. Leg movements and respiration were recorded by kymograph, and heart rate (HR) by a cardiograph. General behavior was observed through a one-way screen.

The problem presented to the animals required that they discriminate three auditory conditional stimuli (csi), each of the three being repeatedly reinforced (paired) with a specific intensity of electrical stimulation applied to the skin of the left foreleg. An audio-oscillator produced the csi, three notes an octave apart and of slightly less intensity than conversational speech, 256 (csa), 512 (csb), and 1,024 (csc) double vibrations. Table

---

[1] Papers on this work were read before the American Physiological Society (2) and the American Psychological Association (3). This work was done in the Pavlovian Laboratory at the Johns Hopkins Hospital and was supported in part by funds from the American Heart Association.

[2] Dr. Roscoe A. Dykman was a Public Health Service Research Fellow of the National Institutes of Mental Health at the time of the work (1950-1951).

[3] The writers wish to acknowledge their appreciation to Dr. James Deese of Johns Hopkins University, and Drs. William Reese and John E. Peters of the University of Arkansas, who read parts of the manuscript and made many valuable suggestions.

Bureau Ednl. [illegible] Research
DAVID [illegible] COLLEGE
Dated

1 shows the electrical unconditional stimuli (USi) arranged in their order of intensity as USa, USb, and USc. The only significant changes made in USa and USb in the last 200 trials were for the dog Whittier. The intensity of USc was increased for every animal. All of the USi had to be increased for Whittier because he was responding less frequently each day.[4]

There were three major phases of training: (*a*) The dogs were habituated to the camera and the Pavlov stand (a wooden frame which permits the dog a limited degree of movement). (*b*) The animals were given the three tones without reinforcement. This is designated as orienting or preconditioning control training, the purpose being to establish basal behavior levels (motor stepping and HR), and to eliminate the "innate" responses to the three tones. Such preliminary training served to overcome the novelty or strangeness of the situation to the animals, and thus reduced the diffuse excitatory phase of the investigative action. (*c*) After these preliminary preparations, the tones were given in a regular series of combinations with the electrical stimuli (USi). The order of presentation of these signals was always the same; i.e., csa-USa, csb-USb, and csc-USc. The csi lasted 5 seconds and the USi 0.6 second. The period from cs onset to US was 4.6 seconds; thus the stimuli overlapped for 0.4 second. A mechanical timer gave a cs-US sequence every two minutes, and three csi with their USi constituted for tabulating purposes one trial run.

During the middle of orienting training, different intensities of the USi were given at the end of two daily sessions to approximate the values that should be used for conditioning. The main reason for giving shocks during orienting training was to be able to initiate conditioning trials without first approximating the USi intensities. USa was adjusted to an intensity which would just give a consistent leg-withdrawal, USc was just below that intensity which would cause an animal to whine, and USb was intermediary between USa and USc in intensity. In the last 200 conditioning trials USc was adjusted to an intensity which would consistently cause an animal to whine and withdraw its whole body as well as the US-leg away from the stimulus. Slight adjustments were made in the USi during the first 30 to 50 conditioning trials, and a similar variable adjustment period followed the changes in the USi after 200 conditioning trials. For all practical purposes these adjustments were negligible and may have been unnecessary.

Following Trial 200 the criterion of conditioning was placed at 75 per cent

---

[4] A paper this long containing technical terms that are used many times requires certain abbreviations. The following have been adopted from those used in the past work of this laboratory (6).

Stimulus designations:

cs—conditional stimulus, csi the plural form.
  csa—tone of 256 double vibrations presented for 5.0 secs.
  csb—tone of 512 double vibrations presented for 5.0 secs.
  csc—tone of 1,024 double vibrations presented for 5.0 secs.
US—unconditional stimulus, USi the plural form.
  USa—a low intensity electrical stimulus p[illegible]sented for 0.6 sec.
  USb—a medium intensity electrical stimulus presented for 0.6 sec.
  USc—a high intensity electrical stimulus presented for 0.6 sec.

Response designations:

cr—conditional reflect, crs the plural form.
  cra—conditional reflex resulting from the reinforcement of csa by USa.
  crb—conditional reflex resulting from the reinforcement of csb by USb.
  crc—conditional reflex resulting from the reinforcement of csc by USc.
UR—unconditional reflex, URs the plural form.
  URa—unconditional reflex produced by USa.
  URb—unconditional reflex produced by USb.
  URc—unconditional reflex produced by USc.
HR—heart rate, HRs the plural form.

In these experiments the UR is the physiological reflex leg withdrawal produced by painful electrical stimulation. The cr is the acquired "duplication" of the UR elicited by the three auditory signals.

TABLE I

MILLIAMPERAGE FOR THE ELECTRICAL USi

Dog	First 200 Trials			Last 200 Trials		
	USa	USb	USc	USa	USb	USc
Isis	2.0	2.5	3.1	2.0	2.5	3.6
Blanket	1.8	2.4	3.1	1.8	2.4	3.6
Whittier	1.8	2.4	3.2	3.1	3.7	4.9
Schnapps	1.6	2.0	2.6	1.4	2.0	3.2
Average	1.8	2.3	3.0	2.1	2.7	3.8

differentiation. To qualify as successful each animal had to give 18 appropriate reactions to the three different tones in a conditioning session consisting of 24 presentations of the tones; i.e., 18 cardiac excitations or motor crs differing in magnitude so that cra < crb, crb < crc, and crc > cra.

Each dog was given 10 trials (30 sequences of tones and their electrical stimuli) per day for the first 190 trials. After this, all conditioning sessions consisted of seven to 10 trials depending on the restlessness of the animal. In the preliminary orienting training, 10 to 15 repetitions of the three tones were given in each session. Kymograph and cardiac records were taken every 50 to 60 trials during conditioning and somewhat less often during the orienting training.

The daily procedure for conditioning was as follows: *(a)* The dog was brought into the experimental room and allowed to play and run free for a few minutes, then offered all the biscuits he would eat. *(b)* He was placed in the Pavlov stand, and the collar was fastened to an overhanging hook. *(c)* Two US electrodes were applied with electrode paste about one inch apart on the shaved left forepaw. Electrical resistance was measured between the electrodes and if the resistance registered over 1,400 ohms, the electrodes were removed. The paste was then reapplied and thoroughly rubbed in, and this usually lowered the resistance. The average daily resistance for all dogs was 950 ohms with a standard deviation of ±250 ohms. The milliamperages (ma.) of the USi were made about equal from day to day by adjusting resistors of three independent electrical generator circuits in relation to the resistance of the animal. The resistors were in series with the animal. *(d)* Conditioning trials were given as described in the above paragraphs. The voltage drop across the series resistors for each US was observed with a cathode ray oscilloscope while the animal was being stimulated. Several readings were taken with the oscilloscope in the early conditioning sessions until a graph for the various relationships could be obtained. Neglecting the impedance effect of tissue, the ma. passing through the animal was calculated by Ohm's law. The resistors in the generator circuits were maintained at high values to diminish the impedance effect. *(e)* The dog was removed from the camera and the left foreleg was washed with warm water to remove the electrode paste. When the electrode paste was not removed the skin resistance increased on subsequent days. The animal was then petted, fed, and allowed to play and run free in the experimental room.

RESULTS

1. *Preconditioning Control Training*

The orienting reflex variously termed investigatory, focusing, questioning, or "what is it" reaction, is the natural response which an animal gives to most novel stimuli (14, 16). It should be distinguished from the startle reflex which is a far more rapid, "tensing" response to traumatic stimuli. In the orienting reflex the dog turns his head slowly toward the source of the novel stimulus (motor reflex), and this movement is associated with certain visceral changes in respiration, HR, and sometimes salivation. The orienting reflex has been extensively studied (16); it is necessary for the purpose of this paper only to state that it was extinguished before the beginning of the conditioning trials.

2. *Comparison of the HR and Motor cr Changes to the Three Auditory Signals with Respect to their Time of Appearance and Differentiation*

A. *The character of the early responses.* In understanding the learning process it is essential to consider very carefully the behavior of animals in *the first few conditioning trials.* The results for the first three conditioning trials are presented below. The average HR and the motor response (leg withdrawal) changes occurring to csi and USi are given on the second and third lines; and the responses occurring in the period before the signals on the italicized line above.

The two figures in the italicized line represent the periods 20-0 and 5-0 seconds before the tone, respectively. The figures immediately following the csi are for the tone period (responses in this period are usually designated as crs), and the two figures following the US are for the periods 0-5 and 0-20 following the onset of the electrical stimulation. The motor response reflects the dual aspects of frequency multiplied by maximum amplitude. For example, if the greatest amplitude of a motor response in a certain unit of time was 2 centimeters on the kymograph record, and if this response occurred three times, it was given a score of 6. HR is entered as beats/minute. All the figures below are the averages for the four dogs.

Trial 1

Control–HR (97, 100); motor response (2.5, 0.51)
(A) csa–HR (102); motor response (0.68)
USa–HR (106, 98); motor response (2.0, 3.1)

Control–HR (93, 104); motor response (2.1, 0.48)
(B) csb–HR (96); motor response (0.58)
USb–HR (101, 95); motor response (4.0, 10.0)

Control–HR (93, 101); motor response (1.5, 1.5,)
(C) csc–HR (98); motor response (0.0)
USc–HR (112, 111); motor response (4.0, 4.5)

Trial 2

Control–HR (102, 111); motor response (2.4, 1.2)
(A) csa–HR (107); motor response (0.55)
USa–HR (112, 108); motor response (4.0, 6.5)

Control–HR (96, 92); motor response (3.5, 0.0)
(B) csb–HR (111); motor response (0.43)
USb–HR (107, 103); motor response (4.6, 6.3)

Control–HR (98, 91); motor response (0.0, 0.0)
(C) csc–HR (103); motor response (0.0)
USc–HR (114, 108); motor response (7.5, 8.8)

Trial 3

Control–HR (103, 102); motor response (3.3, 1.5)
(A) csa–HR (110); motor response (3.7)
USa–HR (103, 94); motor response (4.0, 6.8)

Control–HR (90, 89); motor response (2.5, 0.0)
(B) csb–HR (104); motor response (0.39)
USb–HR (105, 98); motor response (6.0, 7.8)

Control–HR (100, 101); motor response (3.5, 1.9)
(C) csc–HR (106); motor response (0.45)
USc–HR (123, 110); motor response (7.2, 15.8)

In these first three trials a cardiac change to the three auditory signals is more consistent than the motor change. Contrasting the motor responses occurring in the five-second period before the tone with those occurring in the tone period indicates that a motor response is about as likely to occur in a resting period as it is to the specific cs. The HR shows an increase starting with the auditory csb in Trial 2. The results for the four animals were consistent at this time.

These early responses indicate that the first change that occurs in an animal in this type of conditioning situation is an emotional sensitization or mobilization. The diffuse and generalized nature of the early cr is well established in conditioning literature (14). There was a general increase in motor restlessness in these early trials, which paralleled the changes in the resting HR.

The control HR averages before Trial 1 were somewhat above the average values for the last orienting session. These averages were based on four dogs, and all of them had a higher heart rate this day, especially the dog named Schnapps. It was an unseasonably warm day, and this may have accounted for part of the increase.

Table 2 gives cardiac and motor changes to the csi by three blocks of trials: the first two trials, the first 10 trials, and Trials 5-10. In this table, responses to the three csi are pooled. If the HR during a cs was higher than the HR before that signal, it was scored as a response. Any motor responses of the US leg occurring to the csi were scored as motor crs. There were no specific leg flexions in the early training.

This method of scoring should take into account the chance happening of a given response independent of the specific auditory stimulation. An HR change has a random occurrence of about 45 per cent. There are some ties; i.e., the control HR equals the HR to a stimulus. The motor response also has a chance frequency, which varied among the different animals. The frequency of motor stepping in the first 10 trials was tabulated for 30 periods of five seconds duration selected *randomly* midway between the successive tones, and the italicized figures, appearing after the name of each dog in Table 2, give the chance occurrence of a motor response.

When the various frequencies were compared by the chi-square technique (disregarding the correlation coefficients), it was found that the motor and the cardiac changes to tones were significant in three of the four animals. If attention is focused on the last five trials, HR is significant in every dog, and the motor response changes are significant in three dogs. It can not be concluded from these results that a cardiac response has any advantage in frequency over the motor cr in the first 10 trials.

Thus, the analysis of the first three trials shows a clearer predominance of cardiac activity to the auditory signals than motor activity, but by 10 trials the motor component is significant in most animals. One must be careful not to conclude that the autonomic activity reflects the emotional process, and that motor activity reflects the learning of a specific adaptation to the environment. All of the early responses seem to have an emotional significance, and it is only later that response specificity and differentiation emerge.

These motor and autonomic data were analyzed for *concomitance;* i.e., whether a cardiac excitation tends to appear on the same trial with a motor cr, and whether the cardiac change and the motor cr are both absent on the same trial. There is about a 45 per cent probability of obtaining a cardiac excitation accompanying any motor response. The concomitance in the last five trials was well above this theoretical expectancy for the dogs Isis and Blanket, but for the other two dogs the concomitance was considerably below the theoretical expectancy.

In these early trials, there was the beginning of a pattern which was to persist throughout conditioning, the tendency for the motor cr to be always accompanied by an acceleration in HR. It should be clear, however, that the HR change is not merely a sensory stimulation resulting from the motor movement, for it can occur in the absence of any motor movement and is often out of proportion to the magnitude of that movement. Past work in this laboratory has indicated that the HR tends to accelerate to an inhibitory conditional stimulus where there is no motor cr (7, 8).

The HR of the four animals in the last training period for the orienting reflex averaged 91 and 90 beats/minute for control[5] and signal. By the end of the first conditioning session the control HR had increased to an average of 107, and the HR to the tones had increased to an average of 119.

B. *The development of differentiation.* Table 3, based on all dogs, contains an analysis of average HRs for various selected periods of time preceding, during, and following the three tones. The period 0-4 is the first four seconds of the cr period, and the period 4.6-5.5 is the UR period. R is a randomly selected 10-second period mid-way between the successive tones. The first block in Table 4 presents averages for the first 10 conditioning trials (Series 10); the second block relates to a test session of eight trials recorded between Trials 191 and 200 (Series 200); and the third block relates to the criterion test session of eight trials recorded for each animal's best performance, somewhere between Trials 360 and 467 (Series 400).

Table 4 summarizes the average motor results for these same conditioning series. The frequency of the motor cr is summarized in Column 1. The average latency for all motor cr's appears in Column 2. The latency for a single cr is considered as the period from the onset of the tone to the first recorded motor response.[6] Columns 3 and 4 give the average amplitude of the motor crs and URs. The amplitude for any response was taken as the maximum deflection of the kymograph pens both above and below the base line. Columns 5 and 6 give average activity duration (in seconds) for crs and URs. Column 5 is the combined activity for both the UR and the cr, and Column 6 is the UR alone. Column 7 is a summary of "anticipatory" behavior; i.e., the motor movements occurring within a 10-second period prior to the onset of the cs. The analysis here is comparable to that for motor frequency—the percentage of trials in which a leg movement occurred, and it is restricted to the US leg alone. With regard to the "anticipatory" movements, it is interesting that the averages tend to be higher than for the motor movements occurring in periods midway between the successive tones (Table 2), particularly in Series 200 and 400.

These data show that animals can be trained to give different HRs, motor response amplitudes, and motor response latencies to auditory signals representing different intensities of electrical stimulation. The average levels of response to the three tones for all animals were contrasted for each conditioning series by analysis of variance. The cr separation F values to the three tones were not significant in Series 10 for HR, motor latency, and magnitude. Corresponding significance tests for these three measures in Series 200 and 400 indicated that the crs to the three tones were significant with $p < .05$ in Series 200 and $p < .01$ in

---

[5] The term control is used to refer to the 5-second period preceding the presentation of the tones unless otherwise specified.

[6] If learning is acknowledged to be a central process, then in calculating latency it would be necessary to subtract the peripheral latency—that is to say, the latency between the central event and the peripheral manifestation. There is physiological evidence of quicker transmission to striated muscles than over autonomic paths to cardiac muscle. Since the transmission times appear to be but a fragment of the total response latencies, they are not considered.

TABLE 2

FREQUENCY OF THE CARDIAC AND THE MOTOR RESPONSES IN THE FIRST 10 CONDITIONING TRIALS

Animal	% Random Stepping	First 2 Trials—%			First 10 Trials—%			Trials 5–10—%		
		Cardiac crs	Motor crs	Concomitance	Cardiac crs	Motor crs	Concomitance	Cardiac crs	Motor crs	Concomitance
Isis	*47*	67	17	50	70*	73*	70	80*	100*	87
Blanket	*30*	50	50	66	73*	80*	73	93*	93*	87
Whittier	*20*	50	33	50	87*	43*	50	93*	47*	20
Schnapps	*20*	67	17	33	63	7	30	80*	0	20
Average	*29*	59	29	50	73	51	56	87	60	54

* $p < .05$ by χ^2.

Series 400. The between-animal *F* values were significant ($p < .05$) in all three conditioning series.

HR levels differed statistically to the three tones in Series 200 and 400 in terms of the one-second changes occurring during the tone period. Analysis of variance using the individual averages for each animal from second to second showed that HR levels to the three tones and changes from one second to the next during the tone interval were significant ($p < .01$). By Series 400 the pattern of HR changes to tones was well established. In Series 10 there was a tendency for a cardiac acceleration to occur during the first second of a tone. In later series there was a consistent tendency for an HR deceleration to occur during the first second of a tone followed by an acceleration to some maximum value.

Figures 1, 2, and 3 give the changes in differentiation from one recorded conditioning session to the next. Motor and HR differentiation follow a similar time scale.

The HR excitation to the tones showed an interesting phenomenon of gradually decreasing while differentiation increased. The motor crs were irregular: motor crb and motor crc increased during the first 200 trials and tended to decrease in the last 200 trials, and motor cra remained relatively constant throughout (neglecting the temporary changes in performance that were produced by the increases in USi). The fact that a motor cr even occurred to csa may be a reflection of increased sensitivity produced by the USi accompanying csb and csc. In any event, the averages for the motor cr to csa are based on fewer readings than for csb and csc, for the frequency was nearly always less (Table 4). Also, the reverse type of generalization must have been operating; i.e., the USa tending to influence the responses to csb and csc to decrease both their frequency and magnitude. The irregularities in the acquisition curves for the motor cr are very likely a function of the nature of this complex conditioning situation with its complicating emotional features.

Figure 1 shows that HR decreased from one conditioning session to the next except for those changes resulting from the alteration of the US. The rate of decrease is given by the three straight lines in this figure. The disturbance in behavior resulting from the change in the intensities of the USi after Series 200 was disregarded for purposes of this computation. The HR might have

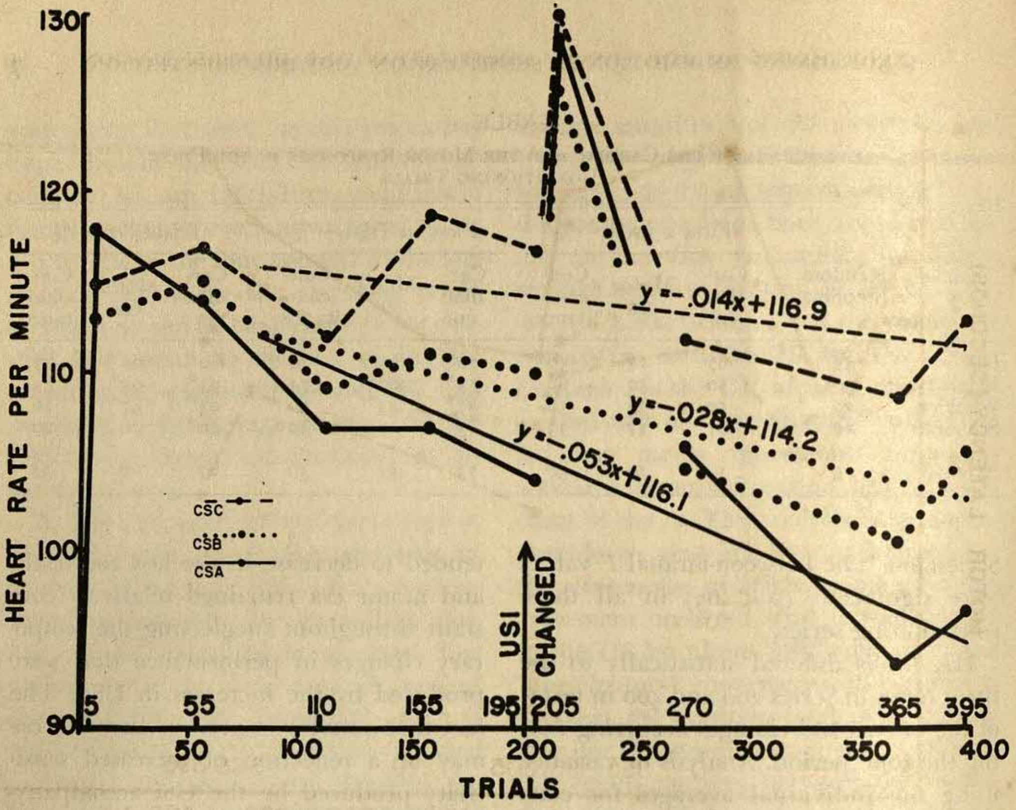


FIG. 1. Development of cardiac differentiation.

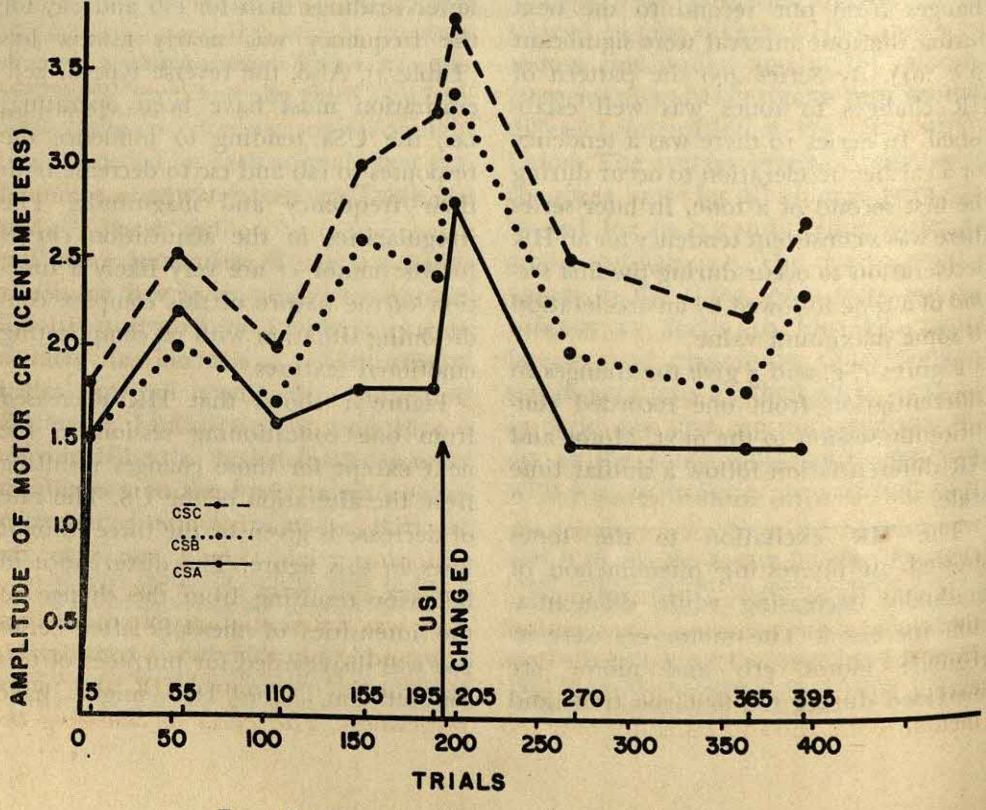


FIG. 2. Development of motor amplitude differentiation.

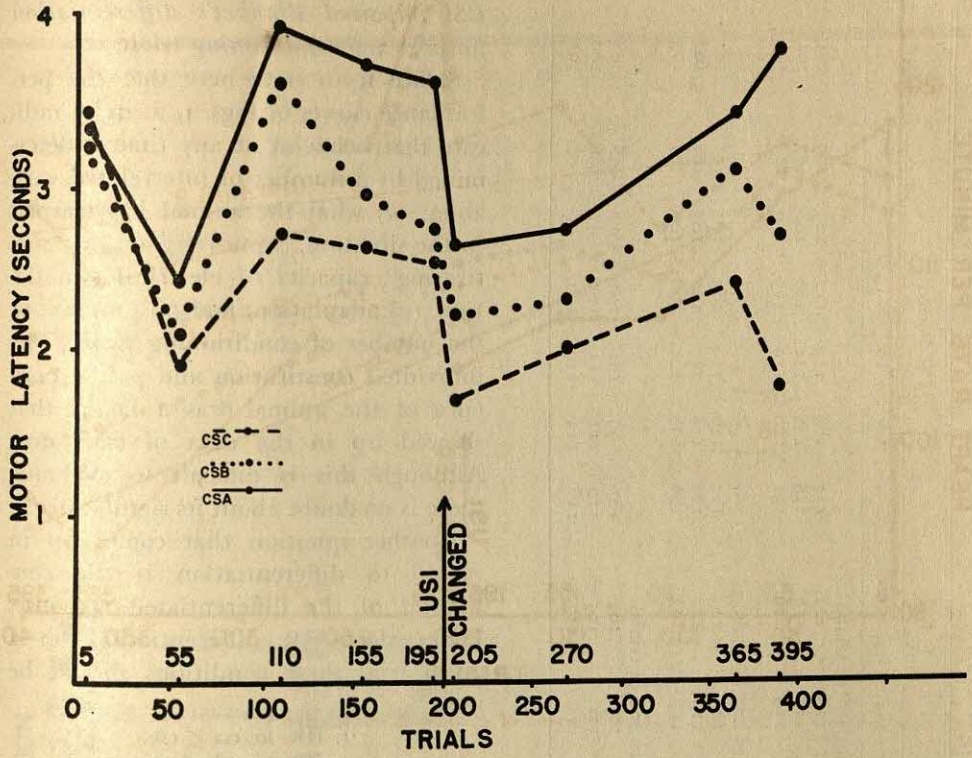


FIG. 3. Development of motor latency differentiation.

Note.—Figures 1, 2, and 3 show the development of differentiation and the effect of changing the USi intensities. Average responses to the three tones for various conditioning sessions covering the first 400 trials are given. The numbers appearing above the abscissa are the "middle" trial for each recorded conditioning session. All conditioning sessions plotted consisted of seven to ten repetitions of the three tones.

reached a lower level had it not been for these changes in the USi.

In general, all "irrelevant" motor movements (any movement other than that of the US leg) decreased with training. Any change in stimulation increased immediately the frequency and duration of random motor movements, such as the first time the tones were given, the first time the electrical stimuli were given, or when the intensity of reinforcement was changed following Series 200.

It may be of interest to note some of the changes in behavior that resulted from the changes in USi. All of the dogs except Isis stopped eating in the experimental room, and this lasted for over 100 conditioning trials. While the animals never appeared to be "fearful" about coming to the experimental room, they manifested symptoms of fear as soon as they entered it. Two of the dogs (Isis and Blanket), who would normally run in and jump up on the stand when called, would not do this following the USi changes. Blanket had to be led into the camera for all the remaining conditioning trials, whereas Isis, after 75 more conditioning trials, would run into the camera and jump up on the stand. Not only was differentiation impaired (Figs. 1, 2, and 3) but there was a considerable increase in the control HR and motor restlessness. *The effect of changing the*

Bureau Ednl. Psy. Research
DAVID [illegible] ING COLLEGE
Dated

TABLE 3

AVERAGE HR BEFORE, DURING, AND AFTER A CS-US SEQUENCE IN THREE SELECTED CONDITIONING SESSIONS
(Time in Seconds)

CS	Pretone Period						Tone Period				UR	UR Period						R*
	25/15	15/10	10/5	5/0	2/1	1/0	0/1	1/2	2/3	3/4	4.6/5.6	6/7	7/8	5/10	10/15	15/20	20/30	
Series 10 (10 trials)																		
csa	102	106	105	108	105	107	110	117	121	128	122	117	108	113	112	109	106	104
csb	105	108	101	102	104	102	108	111	116	122	119	118	110	111	106	108	107	104
csc	108	107	108	112	113	114	114	111	116	119	126	131	116	119	115	110	108	106
Series 200 (8 trials)																		
csa	103	105	106	108	113	104	99	102	110	105	113	114	110	110	103	104	101	101
csb	104	100	102	101	102	104	104	114	114	113	117	123	118	117	109	103	101	104
csc	106	104	105	106	104	104	101	117	120	117	123	126	128	121	108	100	97	105
Series 400 (8 trials)																		
csa	102	100	100	102	104	101	92	95	101	101	104	101	99	101	102	99	98	98
csb	98	98	99	100	104	100	95	105	113	111	105	110	111	111	105	97	96	100
csc	100	100	103	106	110	111	105	114	126	125	117	116	123	119	107	97	97	101

* R is a randomly selected period mid-way between the tones.

USi impaired Blanket's differentiation for over 200 conditioning trials.

Suffice it to state here that the performance curves of Figs. 1, 2, and 3 indicate that behavior at any time is determined by a number of interrelated variables: (*a*) what the animal has learned in the situation, (*b*) novelty, (*c*) the "motivating" capacity of electrical stimulation, (*d*) adaptation, and (*e*) practice, or the number of conditioning trials. The individual constitution and past experience of the animal was a factor that showed up in the data of each dog. Although this is difficult to evaluate, there is no doubt about its significance.

Another question that comes up in regard to differentiation is the *consistency* of the differentiated response. For perfect HR differentiation on a given trial these conditions should be met:

1. HR to csa $<$ csb,
2. HR to csb $<$ csc,
3. HR to csc $>$ csa.

Comparable conditions should hold for motor amplitude and (reverse) for latency. This analysis was carried out for several recorded test sessions; the data are shown in Fig. 4. This figure reflects the consistency of performance from one cs to the next in the various conditioning sessions. The important indication here was that *motor and HR differentiation developed approximately together,* although somewhat greater consistency is shown for motor amplitude than for motor latency and heart rate. The impairment in both motor and HR consistency following Series 200 resulted from the changes in USi intensities.

3. *The Development of a "Time Reflex" Resulting from the Use of Regular Stimulus Intervals*

The time reflex was not as consistent

TABLE 4

AVERAGE MOTOR RESULTS FOR THREE CONDITIONING SERIES

Series	cs	1 % crs	2 Latency (secs.)	3 cr Amplitude (cms.)	4 UR Amplitude (cms.)	5 Total Activity (secs.)	6 UR Activity (sec.)	7 % Anticipatory Activity
10	csa	50	3.3	1.51	2.05	6.9	5.9	45
	csb	55	3.2	1.53	2.26	6.6	5.5	30
	csc	48	3.4	1.83	2.66	8.2	6.9	38
200	csa	68	3.6	1.80	2.40	3.9	2.9	50
	csb	89	2.7	2.40	2.90	4.8	3.2	39
	csc	97	2.5	3.20	3.70	5.3	3.4	43
400	csa	78	3.9	1.60	2.00	4.6	3.9	68
	csb	100	2.6	2.30	3.60	6.6	4.8	45
	csc	100	1.8	2.70	4.40	8.8	6.0	63

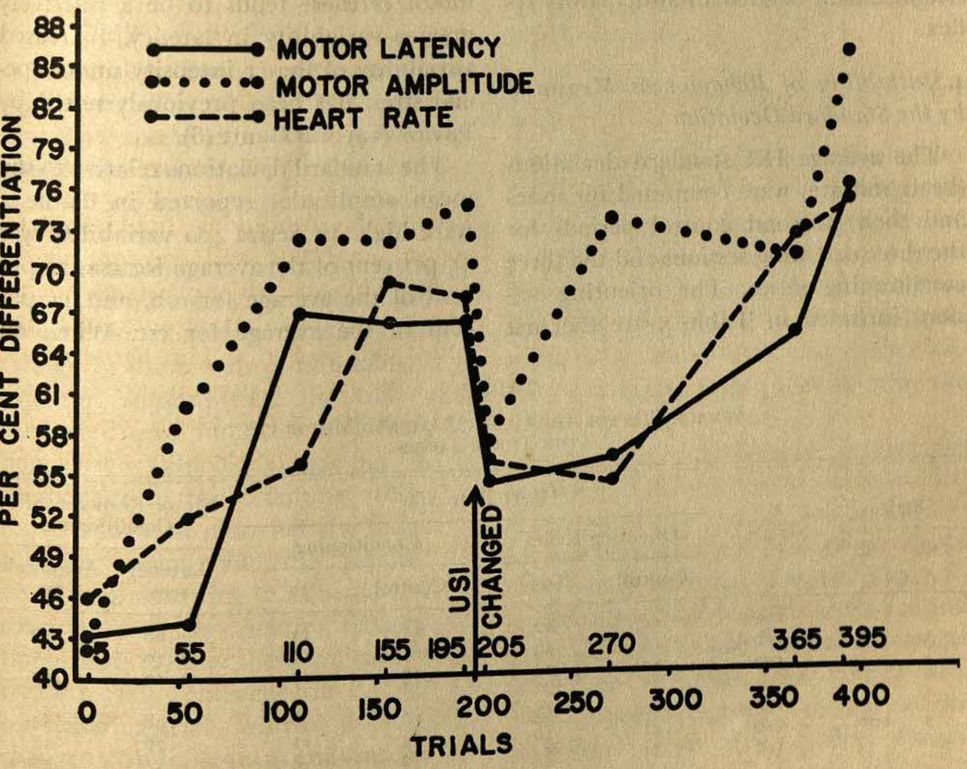


FIG. 4. Differentiation Comparison.

Note.–This figure contrasts HR, motor latency, and motor amplitude differentiation to the three auditory signals at various stages of the conditioning process. Conditioning sessions are the same ones that are shown in previous figures.

in this work as has been found in past studies of this laboratory (7,10). However spontaneous stepping gradually decreased during conditioning (except when stimuli conditions were changed), and all animals showed an increasing tendency to limit their stepping to periods just before the csi. These anticipatory motor flexions were accompanied in two dogs by whining, and one animal (Schnapps) went for several sessions stepping just a second or two before the onset of the conditioning tones. In general, the results were far more variable, and the time reflex took longer to develop than has been characteristic in past work. The complexity of this conditioning situation was probably detrimental to the establishment of such an anticipatory reflex.

4. *Instability of Behavior as Measured by the Standard Deviation*

The average HR standard deviations (beats/minute) were computed for tones and their 5-second control periods for the three orienting sessions and the three conditioning series. The orienting sessions included in Table 5 are the first and last, and one session just before the experiment with electrical stimulation. These orienting sessions are designated in Table 5 as 1, 3, and 2, respectively. This table does not show the changes in variability that followed the changes in the intensities of USi after Series 200. The variability was considerably increased by these changes.

The standard deviations for latencies of the motor crs were considerable in terms of the average values reported in Table 4. For Series 400 the standard deviation was 14 per cent of the average recorded value for csa, 19 per cent of the csb value, and 24 per cent of the csc value. When an animal is responding near the maximum level possible for a motor cr there tends to be a relatively greater variability in latency. Increased variability of the cr intensity under special stress has been previously noted by Pavlov (14) and Gantt (6).

The standard deviations relative to the mean amplitudes reported in Table 4 were high. In Series 400 variability was 45 per cent of the average for csa, 20 per cent of the average for csb, and 11 per cent of the average for csc. When an

TABLE 5
AVERAGE *SD*s FOR THE HEART RATE AND MOTOR CRS TO THE THREE TONES

Series		Tones	Heart Rate *SD*s				Motor Response *SD*s Conditioning	
			Orienting Reflex		Conditioning			
OR	C		Control	Tone	Control	cs	Latency secs.	Amplitude cms.
1	10	csa	11	12	15	14	.98	.50
		csb	12	13	15	15	1.10	.60
		csc	12	14	17	17	1.10	.55
2	200	csa	10	11	10	8	.98	.54
		csb	11	10	13	11	.76	.45
		csc	11	11	10	8	.54	.40
3	400	csa	10	11	14	11	.54	.72
		csb	11	10	16	11	.49	.45
		csc	9	11	15	14	.44	.31

animal is responding to the greater intensities of electrical stimulation, there is less relative variability in amplitude. The average standard deviations for the URs were analyzed for Series 200 and 400, and these results paralleled the amplitude of the cr in direction of change and relative variability.

The standard deviations relative to the mean HRs were also appreciable. The average HR to the tones in the last orienting session was 90, and a standard deviation of 11 is about 12 per cent of this average rate. In conditioning Series 400, a standard deviation of 11 to csb was 10 per cent of the average HR of 106, and a standard deviation of 14 was 12 per cent of the average HR of 117.

TABLE 6

Consistency Percentages for the Heart Rate and Motor crs to the Three Tones

OR	Series	csi	HR		Motor Response Conditioning: Per Cent Flexion
			Orienting: Per Cent Acceleration	Conditioning: Per Cent Acceleration	
1	10	csa	45	75	70
		csb	58	75	65
		csc	48	70	60
2	200	csa	65	46	76
		csb	65	71	85
		csc	80	82	94
3	400	csa	45	41	84
		csb	48	69	92
		csc	45	74	100

5. *Response Consistency*

There were two considerations here: (*a*) the fraction of total HR responses that show acceleration over control rates, and (*b*) the ratio of the total time the leg is flexed or in the process of flexing to the total time the leg is moving. This latter consideration involves two questions: the nature of the early motor responses to auditory signals and the change in these with conditioning. The orienting sessions were the same ones that were used in Table 5. It may be noted from Table 6: (*a*) that the cardiac component of the orienting reflex becomes somewhat more definite following the first session, and this persists for some time, returning to a chance level of consistency as the orienting reflex is extinguished; (*b*) that there is a marked increase in HR consistency in the first 10 conditioning trials, with no appreciable changes with subsequent training for csb and csc, but with a decrease for csa; and (*c*) that there is a gradual improvement in flexion specificity, which might be thought of as the "normal" physiological response to pain.

6. *Comparison of HR and Motor Latency*

An important question is whether the HR response tends to appear before the motor response. The second-by-second HRs for all animals to the various tones in Series 200 and 400 were studied to determine the second in which HR exceeded the basal level, using the one-second average before the tones as the basal level.

The HR to csa exceeded or tied the basal level two to three seconds following the onset of csa, and the csb and csc responses exceeded the basal level one to two seconds following their csi. The average results were similar for Series 200 and 400. Table 3, which contains average values for the four dogs, reflects this. There were numerous instances in the individual trials of a cardiac change occurring within one second following the onset of tones, particularly in Series 10.

TABLE 7

THE INTERRELATIONSHIPS OF RESPONSE VARIABLES

(Series 400)

Y	X	ϵ_{yx}	Significance
cr HR	cr Motor amplitude	.49	$p < .01$
cr HR	cr Motor latency	.57	$p < .01$
cr HR	USi	.62	$p < .01$
cr Motor latency	USi	.62	$p < .01$
cr Motor amplitude	USi	.72	$p < .01$

These data indicate that the average HR latency is less than the average motor latency in the well-established cr (see Table 4). The latency of the HR response is a less precise measure than the latency of motor response. A study of the individual trials showed appreciable inconsistency from animal to animal, and within the same animals as to the precise time a cardiac change would appear following the onset of a given tone.

7. *Interrelations of Conditioning Variables*

Correlation ratios (ϵ) based on the 96 trials given the four dogs in Series 400 were computed to determine the interrelations of conditioning variables. The ratios, which are given in Table 7, show that on a given repetition of a tone-shock combination, the predictability of the absolute magnitude of any component of the cr is low. There tends to be a negative relationship between the prestimulus levels of HR and the stimulus levels of HR, and this factor confounds the ratios that involve HR. It should be apparent that a small change in HR from a high control value may be similar in magnitude to a large change from a low control value.

DISCUSSION AND SUMMARY

1. *Summary of Major Results*

It has long been recognized that the conditioning of a specific reflex, such as the leg-withdrawal reaction, results in the conditioning of the whole organism. The data of this monograph support this principle: the differentiated motor cr is accompanied by a differentiated HR.

This research has shown that the principles that have been developed for salivary and motor conditioning are also applicable to HR. HR is proportional to the intensity of the US; it is generalized to various signals in the early training trials; its differentiation develops gradually as a function of the number of trials; and the use of regular stimulus intervals produces a cardiac time reflex. The relationship of the HR response to the intensity of painful USi had not been previously worked out, although one of the authors had found evidence in a pilot study to suggest that the relationship is proportional (5) and had studied with Peters the effect of graded degrees of muscular tension on HR (15). The parallelism of the cardiac and the motor cr in differentiation was not expected. Past work in this laboratory (8) and theoretical considerations by Mowrer (13) led us to expect that cardiac differentiation would precede motor differentiation.

There was one significant departure from past observed conditioning results: the increased intensities of the shocks did not produce a permanent increase in the magnitude of the conditional motor and cardiac reactions. Figure 1 shows that HR decreased from one conditioning session to the next while the intensities of the reinforcing shocks were relatively constant—both within the first 200 and within the last 200 conditioning trials. The shock intensities, USc in particular, were increased at the end of the first 200 trials, and this had the effect of impairing differentiation and of increasing the magnitude of both motor and cardiac responses. Other interesting behavior changes occurred: animals showed an increased avoidance of the experimental room; three of the four dogs stopped eating in the experimental room and this lasted for over 100 conditioning trials; and motor restlessness was increased. Blanket, who would normally run in and jump up on the conditioning stand, had to be led into the camera for all the remaining trials.

The HR eventually returned to a lower rate than that which existed before the changes in shock intensity, and the motor cr showed a similar but less striking decrement. And all the other behavior changes, except for Blanket's reticence about entering the camera, disappeared before the completion of training. The precipitous drop in HR was completely unexpected, since past conditioning results lead one to believe that an increase in the intensity of the US, once a cr is elaborated, results in a permanent increase in the magnitude of the cr. The decrease in HR was not a reflex response arising from an increased blood pressure. Other work on blood pressure in this laboratory, which we are now preparing for publication, indicates that blood pressure exhibits a similar intersession decrement.

2. *Autokinesis*

When maladaptive behavior develops in the absence of day-to-day stimulation other than that which confronts the animal in the kennels, it is termed autokinesis (6, 8). Autokinesis is said to be present whenever reactions previously restricted to one set of organ systems spreads to involve additional organ systems, or whenever reactions within each separate organ system are greatly intensified; e.g., a mild cardiac acceleration to a tone may become a tachycardia when the animal is introduced to the experimental environment, to the sight of the experimenter, or to stimuli only remotely connected with the original stress conditions (8).

In this research, an *opposite* self-regulative tendency was found. It would appear that autokinesis can be a two-way process—destructive or constructive—and that the direction of the reaction is determined by the nature of the stress conditions and the constitution of the animal. Although novelty may reverse the constructive autokinetic tendency, the novel conditions of this work did not produce an irreversible change.

In explaining the autokinetic tendency of self-regulation, we would postulate that the day-to-day shocks not only reinforced the specific responses to tones but acted in addition as a reinforcement of mildness. Two dogs used as subjects were retested about one month after the end of the experiments reported here. While their differentiation was only moderately impaired, their HR increased in both the stimulus and prestimulus periods. These animals appeared to be far more apprehensive in the experimental environment.

The type of emotional state that noxious stimuli initiate might be termed *apprehensive expectation,* as opposed to the state produced by rewarding stimuli, which might be termed *eager expecta-*

tion. It is suggested that the constructive autokinetic tendency observed here was permitted by a decrease in the level of apprehensive expectation.

3. *Autokinesis and Differentiation*

In this research, the differentiation and self-regulative autokinetic processes tended to be inversely related. This is not always the case. Bridger (1), using a faradic stimulus from an inductorium, found that the HR of his dogs increased on subsequent days, and yet the animals gave perfect differentiation. It is quite possible, however, that autokinetic reactions must fall within certain limits before differentiation can occur, and the study of these limits is an interesting problem for future research. It is suggested that a slow rate of self-regulative autokinesis, or its absence independent of destructive autokinesis, serves as a warning of psychopathology.

4. *Schizokinesis*

Whitehorn (18) and Mowrer (13) have proposed that learning consists first of emotional arousal, and that such arousal, if not too intense, motivates motor, cognitive, or problem-solving behavior. Mowrer describes the emotional arousal system exclusively in terms of autonomic conditioning. Whitehorn has described the emotional arousal thus:

> . . . a biological condition, characterized subjectively as an excited, tense feeling with considerable tendency to act, but with uncertainty as to what to do, and characterized objectively by motor restlessness or activity, not smoothly patterned, with indications of excess effort, as shown in the facial and respiratory musculature, tremor of the voice, and skeleto-muscular action, together with sudden changes in visceral activity. . . . This experience is found, in general, to be unpleasant (18, p. 260).

It has been observed in the past work of this laboratory that the cardiac response often appears before the motor response in conditioning, and that it outlasts the motor response in retention or extinction. Such a split between autonomic and motor responses has been called schizokinesis (6, 8, 9). The evidence for schizokinesis tends to support Mowrer, who has stated, in effect, that a lack of integration (schizokinesis) is the fundamental basis of learning.

The data of the present study offer two ways of evaluating Mowrer's theory: response acquisition and differentiation. Differentiation, as stated before, was characterized by response integration: there was a striking parallelism of HR and motor crs. With regard to response acquisition, the cardiac response had no advantage in frequency of occurrence over the motor cr in the first 10 repetitions of the three tones and their three shocks, but the cardiac response gave some evidence of earlier conditioning when the analysis was restricted to the first three trials. The significance of this latter is highly suspect: all the early motor crs had the appearance of emotional responses; and more important, the HR was electrically recorded, while the motor crs were mechanically recorded. It is thus possible that the early motor responses were missed by our method of recording.

Whether an autonomic response is a better indicator of emotionality than a motor response depends on the subject. Malmo and Shagass (12) have demonstrated that muscle potentials are as good a reflection of emotionality as changes in HR. Also, cardiac acceleration per se is not necessarily an indication of emotional behavior, for while high HRs are obtained by the injection of moderate doses of acetylcholine, animals give no evidence of being emotionally disturbed and these changes in HR cannot be conditioned (17).

In summary, it is believed that the results of the present conditioning study support a dual factor learning theory. Whereas Mowrer identifies autonomic conditioning with the emotional factor, excluding motor responses, we believe that normal, healthy learning is characterized by an *integration* of autonomic and motor responses, at both the emotional and the problem-solving levels. It is suggested that any lack of integration (schizokinesis) occurring at any point in the conditioning process is indicative of psychopathology.

REFERENCES

1. Bridger, W. Effect of mescaline on higher nervous activity. *Amer. J. Psychiat.*, in press.
2. Dykman, R. A., & Gantt, W. H. A comparative study of cardiac and motor conditional responses. *Amer. J. Physiol.*, 1951, 167, 3. (Abstract)
3. Dykman, R. A., & Gantt, W. H. A comparative study of cardiac conditioned responses and motor conditioned responses in controlled "stress" situations. *Amer. Psychologist*, 1951, 6, 263. (Abstract)
4. Gantt, W. H. The nervous secretion of saliva: quantitative studies in the natural unconditioned reflex secretion of parotid saliva. *Amer. J. Physiol.*, 1937, 119, 493-507.
5. Gantt, W. H. Cardiac conditional reflexes to painful stimuli. *Fed. Proc.*, 1942, 1, 28. (Abstract)
6. Gantt, W. H. Experimental basis for neurotic behavior. New York: Hoeber, 1944.
7. Gantt, W. H. Cardiac conditional reflexes to time. *Trans. Amer. Neurol. Ass.*, 1946, 71, 166. (Abstract)
8. Gantt, W. H. Principles of nervous breakdown-schizokinesis and autokinesis. *Ann. N.Y. Acad. Sciences*, 1953, 56, 143-163.
9. Gantt, W. H., & Dykman, R. A. Experimental psychogenic tachycardia. *Amer. J. Physiol.*, 1952, 171, 725. (Abstract)
10. Gantt, W. H., Hoffman, W. C., & Dworkin, S. The cardiac conditional reflex. *17th Int. Physiol. Congr.*, 1947, 15-16. (Abstract)
11. Liddell, H. S. Adaptation on the threshold of intelligence. In J. Romano (Ed.), *Adaptation.* Ithaca: Cornell Univer. Press, 1949.
12. Malmo, R. B., & Shagass, C. Physiologic studies of reaction to stress in anxiety and early schizophrenia. *Psychosom. Med.*, 1949, 11, 9-24.
13. Mowrer, O. H. *Learning theory and personality dynamics.* New York: Ronald, 1950.
14. Pavlov, I. P. *Lectures on conditioned reflexes.* (Trans. by W. H. Gantt.) New York: International Publishers, 1928.
15. Peters, J. E., & Gantt, W. H. Conditioning of human heart rate to graded degrees of muscular tension. *Fed. Proc.*, 1951, 10, 104. (Abstract)
16. Robinson, J., & Gantt, W. H. The orienting reflex (questioning reaction): cardiac, respiratory, salivary and motor components. *Bull. Johns Hopkins Hosp.*, 1947, 80, 231-253.
17. Teitelbaum, H. A., & Gantt, W. H. A method of intravenous injection of drugs from a distance in conditional reflex studies. *Science*, 1951, 113, 603-605.
18. Whitehorn, J. C. Physiological changes in emotional states. *Res. Publ. Ass. nerv. ment. Dis.*, 1939, 19, 256-270.

(Accepted for publication November 27, 1955)

Vol. 70, No. 16 Whole No. 423, 1956

Psychological Monographs: General and Applied

Perceptual Development in a Distorted Room: A Phenomenological Study[1]

MELVIN WEINER

The Menninger Foundation[2]

I. INTRODUCTION

THIS study is concerned with a series of investigations on the development of perception in adults. The literature on developmental problems with adults is scanty, though Ames (1, 2) and Kilpatrick (10) have recently carried out a series of experiments in perceptual learning. Their goal has been to focus on changes in perception brought about by certain controlled laboratory experiences. The experiments by Kilpatrick, utilizing the Ames distorted room, are noteworthy, and have been of help in the design of this investigation. In the field of child psychology, the pioneering work of Piaget (e.g., 15, 16) on developmental problems stands out. His theoretical formulations serve to stimulate experimental studies by giving them a focus and a deeper meaning.

In the present investigation we shall explore such questions as these: How do adults perceptually construct reality? How do adults become more coordinated with their surroundings and develop veridical (i.e., environment-adequate) perception? We designed an "experimental world" where the individual's environment is new and strange to him, and where the application of old habitual ways of acting and perceiving is inadequate. For the individual to adapt his perceptions so that they are coordinated with the new environment that we devise, he has to develop new ways of perceptual functioning and new percepts. He has to learn from his contact and interaction with his new environment, and has to integrate these new experiences into his perceptual functioning. The plan is to obtain a careful, detailed phenomenological record from *S* in his attempts to adjust to this new world through the development and emergence of new percepts.

The "experimental world" consists of a specially designed room which is perceived as a "square" room, even though it is objectively distorted. In addition, when real-life objects and people are placed within this room, the objects and people themselves are perceived as distorted. The room is so designed that the rear wall, instead of being square, with rectangular windows, and in the frontal plane (as in everyday rooms), is trapezoidal in shape with trapezoidal win-

---

[1] Based on a dissertation submitted to the department of psychology of the University of Kansas in partial fulfilment of the requirements for the Ph.D. degree. The experiments were carried out in the department of research of The Menninger Foundation, for which grateful acknowledgment is due The Menninger Foundation and the National Institutes of Health, USPHS, for their support of this research. It is also a pleasure to acknowledge the encouragement and advice of Drs. Alfred L. Baldwin, Fritz Heider, and Gardner Murphy.

[2] Now at the Laboratory of Psychology, University of Geneva, Geneva, Switzerland.

Bureau Ednl. [illegible] Research
DAVID [illegible] ING COLLEGE
Dated ....

dows, and is tilted out of the frontal plane.

Thus we set up a situation where a disequilibrium exists between the individual and his environment. On the one hand, the observer perceives the room as square, while the room is objectively distorted. On the other hand, objects are perceived as distorted to "fit in" with the perceived squareness of the room. There is a lack of correspondence between the appearance of the objects and room and their objective nature in the environment.

What we are concerned with is to attempt to provide *S* with the opportunity to experience new relationships of objects and surroundings, so that he may develop a new perceptual adaptation to the "experimental world" which we have provided for him. Given the opportunity to view stationary and moving objects from varying perspectives, and to have motoric contact with the room by touching various parts of the room with a stick, the questions we first raise are these: How does *S* learn to perceive the room as it objectively is? How does the observer learn to integrate new experiences into his existing mode of perceiving, and thus establish a new adaptive equilibrium freed from distortion?

The next problem is this: What is the nature of the perceptual development? In other words, what has *S* learned? Under what conditions does *S* learn a *specific* perceptual response when a particular stimulus distribution is produced on the retina? Or, under what conditions does *S* learn a *generalized and transposable* environment–adequate perceptual orientation, in which he would be able to perceive the objective nature of various distorted "experimental worlds?" To test the generality of learning, in this part of the research the observer is presented with a differently-shaped distorted room, but having a projection equivalent to that of the first distorted room. The question is: How does *S* perceive this distorted room after having adequately learned to perceive the objective nature of a differently-shaped distorted room? If objects are then introduced, does the individual maintain a constant relationship to object-characteristics? Would this second distorted room then be organized in terms of the stable and enduring characteristics of the objects within it? Or would *S* have to learn again step by step to perceive the constancy of objects and the objective nature of the second room's distortion? If *S* learns a generalized object-framework relationship, we would expect him to carry over a new way of perceiving that would short-cut a tedious relearning process. Objects should then be perceived as invariant, and the second distorted room should be perceived as it really is in a shorter time. A newly-developed generalized way of perceiving should provide for a general adaptive orientation, which should survive multiple transformations, and should be applicable to the new situation of the second distorted room.

II. Apparatus

The experimental distorted room is a specially constructed room which, when viewed monocularly from the viewing point, appears to be cubical in shape, with 4-foot sides and solid right-angled corners and rectangular windows of equal size (see Figs. 1 and 2). Objectively, however, the back wall of the room is distorted; i.e., the room consists of a "normal" cubical section of square side walls, ceiling, and floor, and a "distorted" back wall section. The back wall and windows are trapezoidal in shape and tilted out of the frontal plane (see Figs. 3 and 4).

There were four principal reasons for

designing the kind of distorted room utilized in the present study. 1. We wanted a design that would facilitate methodological clarity in terms of more adequately representing the nature of the change in perception. The previously existing rooms, being designed with the distortion on all sides (the floor, ceiling, left and right walls, and the back wall), create a problem in design and measurement. With a room distorted on all sides the perception of the room does not change as a whole, and consequently, different parts of the room do not change at the same rate or in the same sequence. Since perception change in any particular part of the room is not an index of the total change, but an index of change only in that part, it is insensitive to both the total amount of change and the nature of the developmental sequence of the change. For example, *S* may perceive a slant in the rear wall. But then, instead of a further change in the rear wall, forward or backward, *S* may perceive some slant in the ceiling. Perhaps there follows a perceived slant in the floor. Then *S* may perceive even less slant in the ceiling, and so on. The manifest sequence of development therefore may not truly reflect the actual development, since while it may seem that no change or even a decrease in the slant of the rear wall may reflect a standstill or regression in development, very rapid and important progress may be going on when other parts of the room are considered. The results derived from such a distorted room may lead to conclusions which are vitiated by the inconsistencies in the perceptual changes in the different parts of the room. The room designed for the present study overcomes these difficulties,

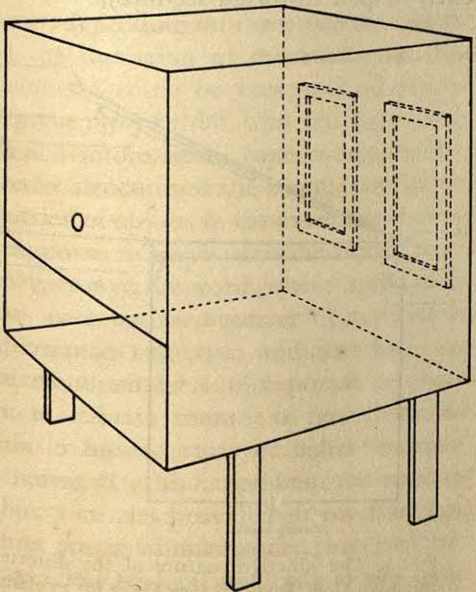


FIG. 1. The phenomenal appearance of the distorted room: perspective view.

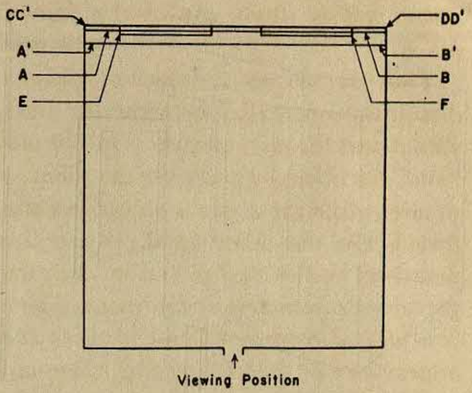


FIG. 2. The phenomenal appearance of the distorted room: plan view.

Explanation of symbols:

A. Left end of movable cable attached to back wall section.
A′. Left end of movable cable attached to cubical section.
B. Right end of movable cable attached to back wall section.
B′. Right end of movable cable attached to cubical section.
C. Upper left corner of back wall where it meets cubical section.
C′. Upper left corner of cubical section.
D. Upper right corner of back wall where it meets (in projection) cubical section.
D′. Upper right corner of cubical section.
E. Upper left corner of left window frame of back wall.
F. Upper right corner of right window frame of back wall.

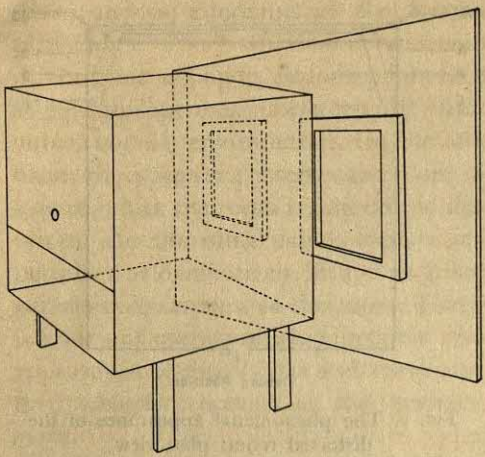


FIG. 3. The objective nature of the distorted room: perspective view, Room 1.

so that more reliable and consistent measurements may be made. By distorting the room in one plane, perceptual change in the distorted plane is a more complete index of the total change, and more directly reflects the sequence and alteration in perception as it occurs. This is not to deny that a room distorted on all planes may be valuable, since investigation of how part-perceptions are integrated is worth while in its own right.

2. We wanted to design a room where overlay cues could be utilized, thus greatly increasing the lifelikeness of the room. The present room is designed to look as if there were a corner where the walls came together, but where objectively there was empty space where the corner was supposed to be. This seems paradoxical, but Fig. 4 indicates that there is a gap between the back wall and the floor, the ceiling and the right side wall.

3. Rooms in everyday life are not just "four walls," but "four walls surrounding objects." Thus, a design was required in which a variety of objects and people could be easily set up and moved about by both *E* and *S*. It is through the gap that objects can be easily introduced and manipulated. In this respect the design of the experiment becomes basically different in focus from previous investigations. In the present experiment, because of the presence of objects in the room we may be able to examine the nature of the relationships between objects and their spatial framework, whereas in previous investigations (10), because *S* was provided with only a stick and rubber balls to be pointed and thrown at various parts of the room, the object-framework aspects of perceptual functioning were not emphasized. Thus, as in the emphasis given by Gibson (6) and Koffka (12), our focus shall be on the interrelationships between the objects-in-space and their spatial framework.

4. Finally, we wanted to design a room which would be relatively simple in construction (the distortion is designed on only one plane, while the main section remains a cube) and flexible in utilization (i.e., the same cubical section can be converted for use with several differently-shaped distorted sections).

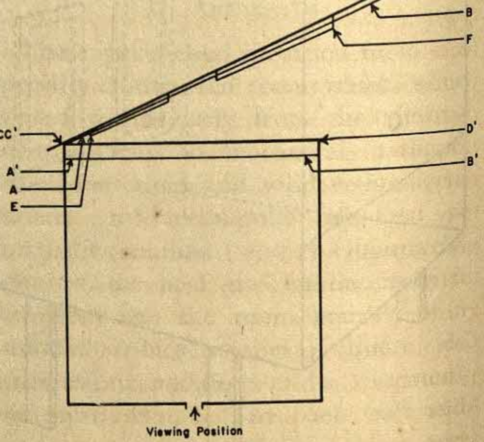


FIG. 4. The objective nature of the distorted room: plan view. Room 1. (See Fig. 2 for explanation of symbols.)

III. EXPERIMENTAL PROCEDURE

A. Experimental Setup

The *S* is brought into the laboratory with a patch over his left eye. He is seated in front of the viewing position of the distorted room, without having been able to see any part of the room or the laboratory. The *S* places his head in an adjustable headrest in such a position that his right eye is at the projection point. He is then asked to open his eyes. Since a patch is over his left eye, he views the room monocularly. After *S* has had an opportunity to observe the room *E* asks, "What do you see?" and follows up *S*'s responses with, "How does it look?" "Can you describe it (more fully)?" "Is there anything else?"

B. Training Procedure

After this initial description of the room is obtained, objects are introduced by *E* to fit the needs of what may be called a training program. Each object is devised and introduced into the room in such a way as to be perceived differently, depending on the objective nature of the distortion of the room. In this manner *S* could be presented with *perceptual information,* and not just intellectual information, concerning the objective structure of the room. The introduction of objects in the experiment does not, however, imply training in the ordinary sense of the word, where the trainee may be a passive observer of a regimen of training activities, and in which he merely attempts to incorporate or integrate a certain amount of new information in his repertory of behavioral adjustments. In this experiment the various objects or cues provided in the training series are not immediately perceived by *S* as cues. In fact, not only is the cue-value of the object not immediately apparent to *S*, but the cues present a conflict for *S* and set up contradictions that have to be resolved. The *S* must, through active participation, actually discover or disclose perceptually what the cue-value is; in other words, he has to learn to utilize the objects as cues. Only in this sense can the objects introduced into the room be considered to involve training. Thus, a variety of everyday objects as well as people are introduced in the experiment to serve as cues, and will be referred to as cue-demonstrations. It was expected that the richness of the objects would not only increase the lifelikeness-character of the experiment, but also increase the potency of the objects as cues.

An example of how a cue-demonstration functions is the following:

In the early stages of *S*'s viewing of the objects in the room, the objects are perceived as distorted (e.g., people seem to walk through solid walls, cigarette packages seem to shrink in size, and marbles seem to roll uphill, etc.), while the room is perceived as perfectly "normal," regular, and lifelike. But after varied experiences with the cue-demonstrations (e.g., *S* is given opportunity to view the cigarette packages moving across the room in various directions), the perception of the objects in the room begins to change, and concomitantly, the objects begin to serve as cues for the objective nature of the room. For example, no longer does the cigarette package look as if it were shrinking in size as it moves across the distorted rear wall. Coordinate with the cigarette package's being perceived as remaining the same size, it comes to serve perceptually as a cue to distance. Then, parts of the rear wall begin to be perceived as they really are; i.e., as distorted, irregular, and not like parts of a real-life room. It should be noted, therefore, that each cue demonstration serves a dual purpose: *first,* that of providing *S* with perceptual cues for the objective nature of the distorted room, and *second,* that of providing *E* with data on the changes in *S*'s perception of the objects *themselves* as they begin to serve as cues.

A list of the cue-demonstrations utilized in the experiment and the frequency with which they were introduced during each hour for one of the *S*'s is presented in Table 1.[3] Since each cue-demonstration exists as an integral part of the results derived from them, a complete description of each cue-demonstration and how the cues were introduced is presented in the section on Results. An attempt was made to present the objects in the room in a sequence of perceptual potency. In other words, the first objects or cues presented to *S* are predominantly *perceptual* in nature and contain a minimum of *intellectual knowledge* or "give away" information. For example, the cigarette package demonstration is presented first and most frequently, since it is very difficult to "intellectualize" this cue and since it is a potent cue perceptually. In contrast, the back wall movement and binocular vision demonstrations are presented later and less frequently.

C. Criteria for Perceptual Change

1. Phenomenological description of the total structure of the room. A description of how the room looked and of the intervening emotional and thought processes of *S* was obtained through the minute-by-minute reporting of the *S*'s experiences. Only after *S* has had opportunity to express spontaneously his style of perceiving, attending, thinking, and feeling about the experiment would *E* attempt to have *S* elaborate on various components that may have been omitted. Even though omission and the nature of omission is significant in itself, it is important to attempt to gain some insight into the facets of behavior that are not easily expressed by *S*. Thus, after such initial questions as, "How does it look?" or "What do you see?" and also after a certain amount of emotional support is given *S*, as the experiment progresses, *E* may ask such questions as: "How do the window frames look? What is their shape and size? How did you go about looking at the windows; the cigarette packages? What are you looking at now? What were you looking at when the cigarette packages were moving? What were you trying to do in pointing the stick at the windows? What were you thinking about when you said that the cigarette packages were shrinking in size? How did you feel when you said that the room looked like it was changing in shape?" etc. The *E* attempts not to give *S* any "leads," but rather to focus attention on various aspects of the situation that may have been omitted.

The *S* is not given much direction as to *what* he should see, or *where* he should look, or as to *how* or in *what language* he should describe it. To obtain the full richness of perception and its development in its most natural manifestation, it was considered expedient to examine the processes of development in *S's* own medium, his own way of operating, and at his own speed.

2. Phenomenological description of objects in the room. A full description of how the cue-demonstrations or objects looked to *S* was obtained as described above.

D. Transfer Experiment

To test the generality of perceptual development, a differently-shaped distorted room is utilized. This transfer

---

[3] The binocular vision demonstration was eliminated for *S*-3. However, essentially similar results were obtained.

TABLE 1

Frequency and Timing of Cue-Demonstrations (*S*-2)

Session Number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	Totals
Length of Session (Hours: Minutes)	1:00	1:40	1:15	0:30	1:30	1:15	0:40	1:05	1:00	1:15	0:40	1:00	1:15	1:10	1:05	1:15	1:00	1:15	0:50	1:40	2:35	24:55
Cumulative Time	0:00	1:00	2:40	3:55	4:25	5:55	7:10	7:50	8:55	9:55	11:10	11:50	12:50	14:10	15:20	16:25	17:40	18:40	19:55	20:45	22:25	24:55
Cue-demonstration:																						
Back wall cigarette package	2	20	8	6	16	8	4	6		4	4	4	8		6	1		4	4	1		106
Designed cards			1	1	1	1	1	2				2						1	1			11
Moving yardstick					1			1			1			1	1	1		1				7
Stick pointing	1	1					1	1				3	1		1	4	2	1		1		17
Man and chair	1, 0*	1, 0	0, 1	0, 1	1, 3	1, 3	1, 1	0, 3	0,3	0, 6	0, 2	0, 4	0, 5	0, 6	0, 5	2, 2	1, 0	1, 4	1, 3	1, 2	4, 1	15, 55
Rectangular card					1	1			3	1	1	1	3	3	2	1		1		1		19
Trapezoidal card						1				5		1	2	2	2			1	1	1	1	17
Back wall movement														1			4					5
Binocular vision		2																			1	3
Outside light					1			1			1		1	1	2	3		1	2	1	2	16
Fluorescent light		3			1	1		1			1	1	1		1	1		1	1	1		14
Light variation	2	3				1		1			1		1		1			1	1	1		13

* Presented in coordination with another demonstration.

Bureau Ednl. Psy. Research
DAVID H.
ING COLLEG
Dated
Acs No

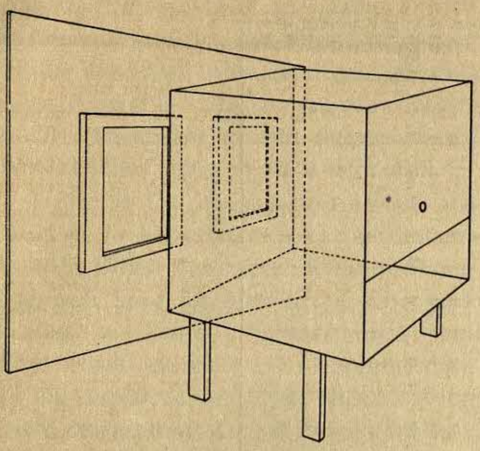


FIG. 5. The objective nature of the distorted room: perspective view, Room 2.

room (which will be referred to as Room 2, as contrasted to the training room, which will be referred to as Room 1) is an equivalent projection of the first distorted room, but distorted in exactly the opposite way. Whereas Room 1 is distorted to the right, Room 2 is distorted to the left, but in precisely the same manner (see Figs. 5 and 6). Since both rooms are projections of a square room, they should initially be perceived as square, such as indicated in Figs. 3 and 4. The rooms were pretested and their perceived squareness was empirically verified.

After many hours of presenting different objects in the room and then repeating the presentations, *S* finally perceives Room 1 and the objects in the room as they really are. Then, without *S*'s knowledge, Room 2 is put into place. The *S* enters the experiment just as if it were a continuation of the training with Room 1. He looks into Room 2, and just as with Room 1, begins the session with a complete phenomenological description of how the room looks to him. Similarly, objects (i.e., the cue-demonstrations) are presented just as they were presented in Room 1. Thus, the same procedures are utilized in Room 2 as in Room 1, but for a different purpose; i.e., to test the generality of learning. Accordingly, phenomenological data on the perceptual changes in the objects and the total structure of the room also are obtained throughout the transfer experiment.

E. Subjects

The data for the experiment were obtained from three female *S*s. All were nonprofessional employees (e.g., secretaries) at The Menninger Foundation. *S*-1 was tested only in Room 1, for 20 hours; *S*-2 in Room 1 and then several months later in Room 2, for a total of 35 hours; *S*-3 in Room 1 and immediately afterwards in Room 2, for a total of 60 hours.

IV. RESULTS

A. The Training Experiment

Because of space limitations, we shall present here the data from only one *S*,

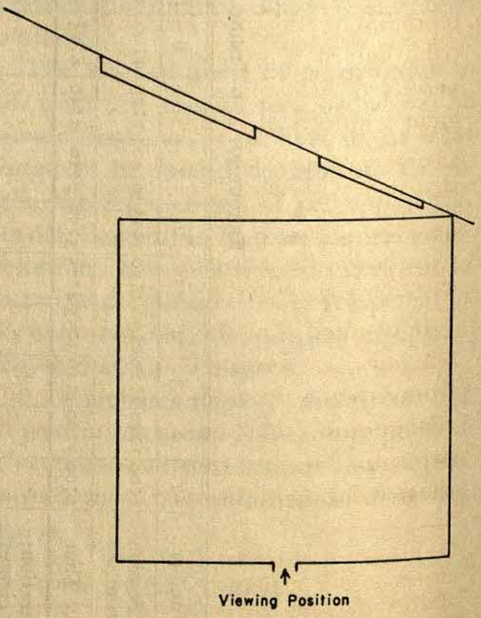


FIG. 6. The objective nature of the distorted room: plan view, Room 2.

namely, *S*-2. We shall later discuss the general trends of development indicated by the three *S*s used in the experiment. After each excerpt or group of excerpts from a protocol, we shall describe in psychological terms the meaning of the raw data. An indication of the number of hours experience up to, but not including the time of the training session being discussed, precedes each excerpt. For example, 2:10 represents two hours and ten minutes of previous experience.[4]

1. Total structure of the room.

0:00 (*S* looks into the room for the first time. There are no objects in the room. The *E* says to *S*, "Open your eyes and look into the room. What do you see? Just tell me how the room looks.") The room looks like a regular room. It's rectangular. The windows just look rectangular.

1:00 (At the beginning of the next hour *S* looks again into the bare room.) I still have a feeling it's square. It looks square. The windows look about the same. They are rectangular.

These initial remarks demonstrate the effectiveness of the experimental room in reproducing the proximal stimulus distribution of a square room; i.e., it is perceived as a square room with rectangular windows of equal size.

1:00 (The *S* has just previously viewed the cigarette package moving into the distance along the back wall.) I wish the windows weren't there. Those darn windows. You get the feeling you're looking at something that doesn't make sense. They are perfectly straight.

It is noted that the perception of the windows as rectangular and in the frontal plane provides a strong stimulus configuration that sets up a strain in the perceptual system, since it conflicts with the perception of distance induced by the perceived movement of the cigarette package into the distance.

1:00 (The *S* views the room with one eye, but had just previously viewed the room with binocular vision.) It's a square room. The illusion is back. I can remember how the distortion is, but I can't see it that way. Up at the ceiling, though, I can really see it going up at an angle. But at the tops of the windows, I know how it is, but it doesn't look that way. The string (on which the cigarette package moves in the cigarette package demonstration) seems to be going down a little and just a little on the bottoms of the windows. I can see the gap; the corners do not meet up any more. I can feel the space in there. (The *S* views the room binocularly again, and then back to monocular vision.) That space seems to close up. The depth, I don't see it. It doesn't look like a real corner, but I can't see the depth. . . . It doesn't look distorted.

Even though *S* had just previously perceived the room as it objectively was with binocular vision, there was no complete carryover to monocular vision. This indicates that an intellectual awareness alone brought about by seeing the room in its objective shape with the aid of binocular cues does not directly alter the mode of perception with monocular cues. However, the beginnings of changes in perception in the direction of veridicality can be noted. At first, there is perception of some space, and concomitantly, the string and the bottoms of the window frames appear slanted.

2:40 (The *S* views the bare room at the beginning of the third session.) It's about the same as yesterday. It still looks square, but not as square as it did. I have no feeling of solidity about the back wall. It is not part of the room as it was in the beginning. It looks even less solid now than it did in the beginning today. I sort of get the feeling that there is an angle to the wall, but I don't see the trapezoidal shape of the windows. The left window looks perfectly rectangular, but the right window does not look perfectly rectangular. The right window looks a little larger.

One of the first feelings that *S* manifests in the early stages in the development of the perception of depth relates to its lack of definiteness and solidity. However, the feeling of definiteness increases with further experience. With the early perception of the angle of orientation, even though this may be derived from a diffuse feeling of depth, one finds the beginning of changes in the perception of size. This demonstrates the relationship between perceived size, shape, and distance.

---

[4] Individual demonstrations were not timed, therefore the indication of time intervals is discrete and only covers the time *up* to but not including the separate parts of each individual session.

2:40 (The *S* has just viewed the cigarette packages moving to the left [L] and right [R].) When I look at the room as a whole, I can look at the different parts of it and see distortion, but I can't look at the whole. For instance, the ceiling, I see a definite angle, a real slant. When I look at the ceiling, then I can see an angle, and I can follow the line of the windows with the ceiling, and then they look as if they are slanted parallel with the ceiling, slanted toward the upper right. The lower right corner throws the thing back into a room. I can't see the space in there, even though it is not a real and solid corner.

There is an indication that an undifferentiated "whole-perception" plays a part in impeding progress in perceptual change. When *S* is able to differentiate the whole into its parts, there is a development toward perceiving the distortion. It is the relating of the parts to the whole which helps to bring about changes in the perception of the window frames.

4:25 (The lower cigarette package is moved from the left part to the right part of the back wall without giving *S* notice that the package is going to be moved.) The wall just seemed to move back when I wasn't expecting the cigarette package to move, but it didn't do it when I was expecting it. (Pause) I just got the overall feelings that I had been shoved back a couple of paces. I got the feeling suddenly that there was more distance between the wall and me. It just seemed to open up. It was a contrast between the way I had been looking and the way it suddenly became.

The perception of distance brought about by the perceived movement of the cigarette package into the distance resulted in the wall's being perceived farther away from *S*'s stationary body. Then after awhile with no change in the stimulus field, *S*'s body was perceived farther away from a stationary wall. This indicates the potency of the framework in resisting change and maintaining as stable a configuration as possible at the expense of the illusory experience of body movement. When there is a stress in the perceptual system, something has to give. In the present case after the increased distance was attributed to movement of the back wall, the framework resisted the change, and part of the distance was attributed to movement of the body.

These results may be compared with those of Witkin, where *S* was seated in an upright, stationary chair while enclosed in a small room which was tilted left and right. Under these conditions the movement of the room with chair stationary induced an illusory experience of body movement. Witkin states:

> As the relation between body and field changed, owing to movement of the room, either some or all of the change was attributed to movement of the body. When *all* the change was attributed to body movement, the room was perceived as completely stationary, even when it had been moved through an arc of 70°. The direction of the illusory body movement was, of course, opposite to the objective movement of the room. When the change was attributed only *partially* to the body, the *S* did perceive movement in the room, but assigned some of the change to movement of his own body. At such times, the room was perceived as moving in one direction—the one it was actually taking—and the body in the opposite direction. Since part of the objective movement was "taken up" by the body, the perceived speed of room movement was less than its actual rate (21).

In the distorted room the perceived distance induced by the movement of the cigarette package remained invariant, whether the increased distance was "made up" by or attributed to the wall's being perceived as farther away from the stationary body, or to *S*'s body being perceived as farther away from the stationary wall. (Cf. also Duncker's experiments in induced motion (5), Koffka (12) for a further general discussion on the role of the framework in perception, and Hilgard (9) as to the significance of the framework in relation to the initial perceptions in a distorted room.)

5:55 (The designed cards are placed on the upper and lower edges of the window frames.) It looks different. It doesn't look like a room. The difference is the space I see in the corner. It doesn't look like a corner. The cigarette package looks so much in the foreground. It looks near and the other one looks farther away. It is nearer. The wall is no longer a part of the room, and I could see it the minute I looked in. The three cards give the lower window a slant. The left window looks rectangular, the right window looks a little wider and it looks as if it slants just a little. Every other time when I first looked in, it was a room, but this time there is a square part, and a back drop.

The perception of more space is developed with the support of the presented size-distance cues (i.e., the cards and the cigarette packages), with a consequent increase in the perception of slant and definiteness to the back wall.

5:55 (The two cigarette packages are placed in a stationary position on the right side of the room; the lower one on the back wall at B; the upper one at the right side of the cubical section at B'.) One (package) is closer than it ever was. The close one, and the other one is farther away. I don't think the room is straight at all. The right side looks shorter. . . . The room looks different, because the front end, the side is pulled in. The room is crooked. It's pulled away from the back wall. I can see the space in there. The back wall looks slanted a little, but it's the rest of the room that makes up for the space.

7:10 (The *S* views the bare room at the beginning of the seventh session.) Both the back wall and the room are at an angle. The room is pulled back a little on the right and the wall is a little farther away on the right.

Again, when a strain is placed on the perceptual system (the greater perception of depth), it is not the back wall that completely offsets the stress. This time part of the cubical section itself (the right side wall) is perceived as shorter to "make up for the space." Part of the perception of increased space is attributed to an increased angle of orientation of the back wall and part is attributed to a shrinking of the right side wall. It appears that the perceived squareness of the back wall with its rectangular window frames is a very stable configuration that strongly resists changes.

7:50 (A fluorescent light outside the room is turned on and shadows are thrown onto the back wall.) The windows don't look alike any more. The one on the right looks bigger and definitely slants away from me. I can see the window going away from me on a slant. Before, the windows seemed straight across and the wall, too. Before, I really didn't see space. I just knew it wasn't joined. But I could see the space between the cigarette packages, in certain places. The left window still looks rectangular. The right side wall still looks shorter.

The shadows on the back wall provide cues for the distance and angle of orientation of the back wall. Then the perception of the orientation of the back wall influences the perceived size of the windows.

7:50 The *S* commented that when she would attempt to "hold on to" the trapezoidal shape of the window, she would feel a muscle strain in her eyes, as if she were crossing her eyes. She also mentioned some feeling of slight dizziness and uneasiness. The *S* reported that she missed a day of testing because her eyestrain was quite noticeable; it lasted even after the experimental session. She said that she would not cross her eyes, even if it were for the experiment. It was easier, she reported, not to see the trapezoidal shape of the windows, because then she would not feel this discomfort and tension. She also stated that the right side of the back wall was difficult to get in focus, especially after focusing on the left window, and that it would blur. There seemed to be some kind of "block" in her attempting to focus, as if there were something hindering her focusing on the right side of the back wall: "The right side of the right window frame becomes indistinct, right where I would see the difference (in trapezoidalness)." It was at these times that the eyestrain seemed to have developed. During an interview *S* related that she would often cross her eyes during childhood as a prank. Someone told her that her eyes might remain crossed, and that she would have to have an eye operation and her eyes would have to be put back into place surgically. The *S* expressed a desire to discontinue the experiment if the eyestrain should continue in further sessions. The *S* was given an opportunity to explore her feelings on various subjects; e.g., her childhood learning experiences, her being a subject, the experiment, the relation of the experiment to psychotherapy, her feeling free during the experiment to say what she wanted. In the following session, *S* reported that the eyestrain was gradually decreasing and finally disappeared, and the experiment continued without any further strain reported.

These comments give us some clues as to the motivational dynamics in perceptual development. It seems that learning even in a "cold" laboratory setting is intimately related to the personal life of *S*. There was resistance to learning or changing, especially when it evoked negative experiences associated with childhood. This was manifested through the appearance of anxiety and tension and was acted out through resistance to the experiments. Since it involved less anxiety, it was easier not to learn or change. If progress were to take place, the resistance had to be resolved and motivation to perceive reality had to overcome the motivation not to learn. When the resistance was worked through or resolved, the learning process was continued and progress was resumed.

Part of the perceptual basis for the eyestrain may be attributed to the conflict between cues, i.e., between visual accommodation that would be adequate for a square room and accommodation for a distorted room. In other words, *S* accommodates for equal distance between left and right sides of the back wall, while veridical perception would necessitate accommodating a greater distance to the right side of the back wall.

8:55 (The cards are placed only on the tops of the window frames.) When I first opened my eyes, it looked farther away than I expected it to be. Then it sort of came closer, but it still seemed to be slanted quite a bit. The tops of the windows look slanted with the cards, rather than the cards look slanted with the tops of the window. The bottoms look pretty straight. The windows no longer look the same size to me. The right one looks larger. Just slightly, the left one looks trapezoidal, when I look at the whole thing. It's only. when I look at the tops alone that I can see the slant more.

The influence of the parts (the cards) on the perceived structure of the framework (the window frames) is noted; i.e., the cards are perceived as slanted, and the top part of the window frames where the cards are placed is influenced by this slant. It is when *S* perceives the tops of the window frames independently of the total structure of the room, that vertical perception is facilitated. Since the cards were not presented on the bottoms of the window frames, there was no cue support there and the window frames were perceived as straight on the bottoms.

14:10 (The *S* views the bare room at the beginning of Session 14.) I had to look at it for awhile. But after I looked at it for awhile, the left window looked trapezoidal and the right window frame just a little. I always think I'm going to see it the way I think it is, but when I look in, it isn't.

15:20 (The *S* views the bare room at the beginning of session 15.) Hmm, it didn't look quite so square. . . . The left window looks quite a lot like a trapezoid. The right window doesn't really look rectangular, but I have to look at it to feel it as a trapezoid. After I look at it for awhile, it looks trapezoidal.

Developmental processes appear to take place not only over long periods of time, but even within a single viewing period. For example, after *S* concentrates her efforts on the window frames for awhile, they take on the trapezoidal shape.

15:20 (In the previous demonstration a large trapezoidal-shaped card was placed in front of both windows in frontal and slanted orientations.) When I look all around, and at the other window, the right window looks more trapezoidal. But it doesn't look like it would if it weren't slanted away from me. But I can't think of it as a rectangle, because I can see the slant.

At this level of development the role of the comparison of parts in facilitating the development of veridical perception is indicated.

16:25 (The *S* views the bare room at the beginning of Session 16.) I can really see a slant, and see a difference in the windows—the right one is larger, faintly trapezoidal—without any particular effort on my part. The left one looks trapezoidal in about the same degree as the right one.

18:40 (The *S* views the bare room at the beginning of Session 18.) It doesn't look flat. The wall looks slanted. I can't quite decide how the windows do look. I feel a little lost without the cards on the top. It was easy for them to outline the shape of the windows. I can see the trapezoidal shape a little, but it requires more effort without the cards. The bottoms were easy to see the trapezoidal shape because the cards outline the window frame (on the bottom).

19:55 (The *S* views the bare room at the beginning of Session 19.) Something's missing. It's kind of bare without the upper cards. It looks kind of close. The windows look kind of trapezoidal. (Pause.) It doesn't look so close now. The right window looks larger. It looks different now than when I first looked at the room. The window looks more trapezoidal and I can see more space on the right side.

At this state of development the percept is not fully stabilized; without the support of the size and shape cues afforded by the cards, there is at first a regression in the perceptual outcome. However, when more effort is expended, an advanced level of achievement is attained.

20:45 (The *E* sits on a chair which is placed in the space and *S* is instructed to point the stick at various parts of the back wall.) I can make a pretty good guess (of pointing the stick at the back wall), but I'm not really confident. The chair with you in it gives it more reality.

The effect of the perceptual system on the motor system is noted here. The *S* has made much progress in perceiving the distance and the angle of orientation of the distorted back wall. This is reflected in "good guesses" in judging the distance with the stick. The potent overlay cues provided by the chair in the space help to provide a sense of reality to the space. (Cf. Michotte's work on the phenomenology of reality [13].)

22:25 (Monocular viewing after viewing the room binocularly.) I don't see much difference, but it seems there's less space. The windows look the same. It isn't quite as clear and sharp as with two eyes, but I didn't have any sense of changing.

There is not much difference in the perceptual achievement of monocular and binocular vision at advanced levels of development. Trapezoidalness of the window frames is the same. However, there is less perceived space and the distance is less well-defined.

22:25 (The *S* was asked whether she could possibly see the bare room as she had previously, i.e., square with rectangular window frames.) It doesn't look like a (square) room. I'd have to stop and think about it in order to see it as a room. I'd have to wipe out the way I've been seeing it all along. I can't quite close up that space over there. It doesn't come. Now I can't see them as rectangles. There is something in between that keeps me from seeing them as rectangles both at the same time. I can see part of it as a rectangle, but one corner bulges out somewhere. When I look from one window to another, they look like trapezoids. If I look in one spot, I can see one of them as rectangular, but when I move, then immediately the trapezoidal lines jump up. If I look in, I maybe can see both windows as the same size, but when I look back and forth, I lose it. They are different in size. I can't completely shut it out. The trapezoidal shapes keep jumping out, the corners and the sides are trapezoidal. . . . In this situation I felt free to learn it the way I preferred, and not forced to learn it in any particular way. I felt kind of confident that I could learn it in any way I choose. In other situations, I feel forced—that I have to learn things when I didn't want to or like it particularly. (Pause.) I can almost see it as a corner, but I can't accept it. I can't talk myself into my seeing it as a corner. When I look at the rest of the back wall, then it definitely doesn't look like a corner. When I look at the back wall and the whole thing, I get more sense of space than when I just look down at the corner. It just falls in, the slant and shape and size of the windows and everything. It doesn't look like a room. (Laughing.) I'm not going to have to learn to see this as a room again, am I?

At advanced levels of development the percept becomes stabilized and takes on the appearance of the objectively-shaped distorted room. It is noted that it is an active comparing of different parts of the room (e.g., the window frames) and relating the parts to the back wall and the room as a whole that helps to bring about veridical perception. The different parts of the room become integrated among themselves and with the whole room. "It just falls in, the slant and shape and size of the windows and everything." That lack of integration or isolation of parts hinders veridical perception is indicated when *S* reports that she can see a part of the window frame as rectangular when she concentrates *only* on that part. Also noteworthy is *how S* went about learning a new mode of perceiving. She reported that she felt free to learn it the way she preferred. By providing a free and accepting atmosphere for *S* during the experiment, there appears to be less possibility that we have unwittingly imposed on the data any preconceived notions of how perceptual development should manifest itself.

2. *Cue-demonstrations. a. Cigarette package demonstration.* In this demonstration a cigarette package is strung along the back wall from the point A to point B (see Figs. 3 and 4). The cable is moved by a reversible direction motor at the rate of 7 inches per second. Another cigarette package may be moved across the cubical section from point A′ to B′ in the frontal plane at the same speed as the other cigarette package. Each package may be moved at the same time as the other package or independently of it. The package that runs along the back wall is placed on a string positioned below the string that runs across the cubical section. Therefore, the back wall package is referred to as the lower package, while the cubical section package is referred to as the upper package.

0:00 (The back wall [lower] package is moved from the right side to the left side of the back wall [B to A].) The cigarette just grows and goes straight across the room. (The lower package is moved from the left side to the right side of the back wall [A to B].) Now it's shrinking, the cigarette package, it's shrinking.

The cigarette package is perceived as moving from A′ to B′ and B′ to A′, while objectively it is moving from A to B and B to A. The package is perceived as moving in the frontal plane along the back wall of a square room, and consequently is perceived as being at an equal distance from *S* on the left and right sides of the room. Size constancy breaks down and size is perceived as varying with retinal angle. The changing retinal image of the package is attributed to a change in objective size, and not to a change in distance.

1:00 (Before the present demonstration had taken place *S* had viewed a man walking into the room into the space—see man and chair demonstration; had practiced pointing a stick at the back wall—see stick-pointing demonstration; illumination had been varied up and down and thus varied the brightness of the back wall—see light variation demonstration; and had practiced viewing the cigarette package moving to the right [R], to the left [L], and again R and L. Just preceding this report the upper and lower packages were again moved R and L across the cubical section and back wall respectively. Thus, the reported percepts represent an end-result of intervening experiences in the room.) The top one looks larger. That's because it's closer. The other one went down along the back wall. If the windows weren't there, I could see the cigarette package going down and back, because when I keep my eye just on the lower cigarette package, I can see the cigarette package going down and back; then if I look up at the windows, that feeling is gone. Then, it's just a cigarette package that's just changing in size.

When *S* perceives the cigarette package independently of the framework of the room, she sees it as moving back into the distance and at an angle, as it objectively does. But when *S* looks at the window frames and comes under the influence of the framework of the room, *S* perceives the distance to the left and right sides of the back wall as equal. The varying distance is not taken into account, so that perceived size varies with retinal angle.

1:00 (The lower package is moved R and L.) I can see it as going up and down. If I look at it and not think, it looks like it goes straight across and gets smaller. Then it goes down and back, but stays the same size.

If *S* maintains a passive, naive attitude, the perception of size variation of the package is the outcome. However, more active participation brings about the perception of distance and size constancy. The achievement of constancy, however, is not stable at this level of development, as the following excerpts indicate.

2:40 (The lower package is moved R, L, R, L, R, L along the back wall, while the upper package is placed on the right side of the cubical section at B.) When they are both in the right corner, I have a real feeling that there is space in the gap. Then I can bat an eye and everything is changed, changed back into a room. When I keep my eye on the lower cigarette package it seems like the wall changes and becomes out of line and the cigarette package is going away and down. But then when I take my eyes off the cigarette package, there is some magic trick going on and it becomes larger or smaller.

3:55 (Both packages are in a stationary position on the right side.) At first glance I saw a big cigarette package and a little cigarette package, and I didn't think of one as being farther away than the other. Just looked like one was bigger than the other. Now it seems that one is farther away than the other.

The *S* gradually develops a percept in which changes in perspective or retinal size are attributed to changes in distance. In other words, increases in perspective size are not attributed to increases in the size of the package itself, but to the increasing nearness of the package. Perspective size of the object serves as a cue to distance. When this occurs, the perception of the package influences the perception of the back wall. The *S* perceives distance and space between the cubical section and the back wall and the back wall is perceived "out of line." However, the perceptual achievement at this level of development is unstable; the room again changes into a square room and size invariancy of the package is lost. The room has a strong tendency to maintain its squareness and consequently distorts the objects in the room. Perceiving the package independently of the back wall and yet taking it into account by relating it to the back wall facilitates the development of veridical perception of both the package and the back wall.

3:55 (The lower package is moved to the right along the back wall and is not stopped at B, but goes out to the right beyond *S*'s view. In a similar manner the package is moved into view to the L from a position beyond B. This sequence is repeated R and L.) When you move it beyond the corner, it takes away the corner. It goes on with the back wall and beyond the right wall.

When the package is moved beyond point B where it disappears from view, it serves as an overlay cue; i.e., when the package cuts through the perceived corner of the room, it decreases the potency of the apparent solidity of the corner.

4:25 (The upper and lower packages are moved L and R.) The cigarette package does not change in size when I watch it move. It goes away, but it doesn't get smaller; it stays the same. But when it stands still, it looks like a smaller package, even though it doesn't look like it is directly below the upper package. When you move it, then it is different.

The role of movement of the package in increasing its potency as a cue to distance and in helping the package to maintain its perceived constancy is demonstrated.

7:50 (The lower and upper packages are moved L and R.) It makes space.

12:50 (The package is moved L, R, L, R.) When

I kind of look at the back wall, I can tell the downhill movement better than when I look at the cigarette package directly. I sort of look at the back wall without really seeing it. When I look at the package, I can tell that it is going farther away better than I can tell it is going downhill.

At this level of development perceiving the package in relation to other parts of the back wall emphasizes the perceived slant of the package. Formerly, the package could not be adequately related to the framework because the package itself was influenced by the framework in the sense of losing its individuality. Presently, however, the package having been adequately differentiated from its surroundings, it can be related to the framework without being enveloped by it.

19:55 (Upper and lower packages are moved L, R, L, R.) It looks a little more slanted than the bottom of the window. Actually the cigarette package looks parallel to the bottoms of the windows, but when the package moved it seemed more slanted than the window. It looks as if the angle between the windows and the string widens a little as it goes out. When I look at the lines alone without looking at the windows or the back wall or the cigarette packages on the string, it looks like parallel lines. The cigarette packages do something for the lines. The lower one is so much farther away. The difference in the cigarette packages has something to do with shoving the wall out.

After approximately 20 hours the slant of the package is perceived as it is objectively; i.e., more slanted than the window frames. The influence of the package on the perception of the slant of the string and on the angle of the back wall is noted. Both the movement of the package and its relating to and comparison with parts of the back wall help in facilitating veridical perception.

In connection with the cigarette package demonstration we should describe a rather striking effect that was obtained with *S*-2 during the post-training "satiation" series. (See Section IV, B, 1.) Five months after *S* had learned to perceive the room as it objectively was, *S* was asked to come back for further testing. When *S* looked into the room after five months a noticeable regression in the newly-developed percepts was manifested. Therefore, further training was undertaken in which the cigarette package demonstration was again utilized, this time in coordination with a new cue-demonstration—the projecting strips demonstration. The projecting strips demonstration consists of taping four strips of poster board, one inch wide and three feet long, on the ceiling, floor, and right side of the cubical section, respectively, in such a way that they project out from the cubical section until they touch the distorted back wall. This provides a potent distance cue for *S* and has the effect of "pushing the wall back into the distance." With the cue-support provided by the strips, *S* perceived the back wall at its objective angle of orientation and the moving cigarette package as maintaining a constant size. With the lower cigarette package remaining stationary on the right side of the back wall, *E* quickly removed the strips from all parts of the room. The following is a verbatim report of what occurred.

E: How does it look as I take the strips off?

S: Umm, it seems a little closer in a way. The difference is in the cigarette packages. The back cigarette package became smaller. The wall got closer.

E: The back cigarette package became smaller. In what way?

S: It shrank.

E: It shrank.

S: Um-hmm. That doesn't make sense. It means just that the cigarette package shrank. cigarette package logically should have looked larger, shouldn't it? But it shrank—I mean it Because actually if the wall came closer, the became smaller to my eyes, so that means it shrank.

E: Well, just describe what it really looked like. I don't mean that it must make sense. Just describe what you saw.

S: Well, the wall came closer, the space narrower, and the cigarette package became smaller—it became smaller as I looked at it. There was a movement that caught my eye, even though I wasn't even looking at the cigarette package. That movement really came so it looked smaller.

E: You saw the cigarette package grow smaller as the wall came nearer.

S: Yeah, but it was really a movement—you know, such as a movement catches your eye.

E: Were you looking at the wall over here?

S: I wasn't looking at anything particularly. I was just watching you peel those things off. And there it was, really a movement which attracted my attention. Just like I was looking at nothing and I saw something move, so I wasn't watching those things, my attention was attracted to them.

This is a unique form of phi phenomenon; i.e., a perceived movement in the third dimension with no change in the stimulus object. When the distance cues were suddenly taken away (i.e., when the strips were removed), the back wall was perceived as nearer and the cigarette package as smaller. Since size constancy depended on perceiving the object-at-a-given-distance, when the perceived distance changed (i.e., shrank), there had to be a concomitant change in perceived size (i.e., the package also had to shrink). Just as previously there was a developmental progression from retinal size to objective size through the development of distance perception, so in reverse there was a breakdown in size constancy with a consequent regression toward retinal size when distance cues were reduced. These results again indicate that an object is perceived in terms of its localization in the spatial context; when the perception of the spatial framework changes, so also does the perception of the object change.

b. Designed cards demonstration. In this demonstration, 2¼ in. by 3½ in. white cards with black designs printed on them are used. Cards with a square are placed on the top edges of the back wall window frames so that the cards lean against the top edges and back wall for support, or are taped onto the lower parts of the window frames; six on the left window frame and ten on the right window frame. In another part of the demonstration cards with plus designs are taped to the sides of the window frames; seven on the left side of the left window frame and nine on the right side of the right window frame.

At first the cards introduced into the room are perceived as distorted. The whole influences the parts. The cards are perceived as "crooked" or diamond-shaped to fit in with the perceived rectangularity of the window frames. However, when more cards are introduced and placed along the tops of the window frames, they are perceived as receding and on a slant. The parts then influence the whole; "It makes the windows go up on a slant" and the "right-hand window frame looks much bigger now because I can see more cards up there. I saw you put the cards up there and I compared them with you, so I know they are all the same size cards. . . . It didn't seem like crooked cards any more." The comparison of the perspective size and shape of the cards with the size and shape of *E*'s hand helps to bring about a change in the perception of the cards. It is the relating of the cards to something of known dimensions that facilitates the development of constancy of the cards. Since the cards were not placed on the bottoms of the window frames, there was no perception of slant on the bottoms. In the early stages of development the perception of objects in the room has an influence only in the area immediately surrounding them.

With further experience the perspective size of the cards serves as a cue to distance; i.e., the perspective smallness of the cards on the right side is attributed to the receding and orientation of the cards and not attributed to the smallness of the cards. They have, therefore, a constant size-at-a-given-distance. At this level of development, there is not yet a stabilized relationship of perspective shape with perceived slant and at times the cards are perceived as if placed horizontally. However, as the perception of the slant of the cards increases, it influences the shape of the window frames. At advanced stages of development, the perception of the windows develops increasing stability and gains freedom from the cue-support provided by the cards. The *S* states, "I can see the windows as trapezoidal plainly."

c. Yardstick demonstration. In this demonstration a yardstick with large black numerals is held vertically in *E*'s hand and is moved from B to A and A to B at the level of the window frames. The yardstick is held approximately one-quarter inch from the back wall as it is moved along. In another part of the demonstration the yardstick is moved in a similar manner along the back wall, but held in a horizontal position in front of the center of the window frames.

4:25 (The *E* moves the vertical ruler to the left [L], to the right [R] and then L and R again.) When I concentrate on the ruler, it first changed in size. It looked larger and smaller, and then as I looked at the ruler, it stayed the same size, and the background changed. It just sort of receded, and I saw the ruler move forward and backward. When I concentrate on the background, the left window grows smaller, and the right window was larger. It was strikingly noticeable, because of the size of the ruler in relation to the size of the window.

Initially, the perception of the size of the ruler is distorted and changes in size as it is moved along the back wall. Then, as *S* takes account of the distance, the ruler remains the same size and parts of the back wall change (e.g., the size of the window frames change). It is when *S* perceives the ruler in relation to the framework that changes are perceived in the framework itself.

7:50 (The ruler is moved L and R.) They looked trapezoidal while the ruler was passing. (The ruler is momentarily taken out of the room.) Now they don't look trapezoidal any more. When you hold the ruler still, I have to compare the length of the ruler with the length of the other side of the window. And then it looks trapezoidal. But then it sort of fades away when I look at the whole thing. When you move it along, I can compare it as it moves so it is easier to compare.

When *S* compares the object with the background and thus relates the object to the background, size constancy of the ruler is obtained, and the windows change in size; i.e., they also achieve veridicality. However, without the support of the size-distance cue from the yardstick, the window changes back into a rectangular shape.

14:10 (The ruler is moved L and R.) Nothing happened to the ruler when it came in, but it got smaller when it went out. (The ruler is moved L and R again.) Now, nothing happened to the ruler but the window widened. The ruler stayed the same size. I was looking at the ruler, but I didn't shut out the background. I was just looking. I was following the ruler with my eyes, but I was also sort of, I wanted to see everything else, too, so I wasn't looking just at the ruler. I wasn't concentrating on anything, but I was looking at the ruler. I was sort of looking with my eyes and waiting to see what happened. I used to concentrate very hard on one object or on one little portion. I used to shut out everything except this one portion, but I didn't really see the rest of the room at all, wouldn't see anything except this one little thing. And now I can see so many more things and don't do that. I just look in and let everything register, and sort of take the whole thing in. If I looked at the whole thing then (before), I didn't learn as much. I have to feel the relationship of the one little object, fit it in with the room, and see the whole thing. It's just looking in, but not really looking at either one of them. I can see the difference in shape. It's just gazing, sort of.

The *S* expresses the kind of perceptual functioning that leads to a given perceptual achievement. At a primitive level of perceptual functioning objects are not adequately related to the background, because the background is perceived in a global way. "(7:50) But then it (the trapezoidalness of the windows) sort of fades away when I look at the *whole thing*. (14:10) If I looked at the *whole* then (before), I didn't learn as much." In the early stages of development the object is embedded in the background and is influenced by it. Objects are distorted under the influence of the background. For example, the yardstick changes in size as it moves toward and away from *S*. However, taking into account a total differentiated field, both the space and the distance, leads to veridical perception of both objects and the back wall. In other words, higher levels of perceptual functioning allow for the object to be differentiated from the background and to be perceived independently of the surrounding framework, independently enough so that is can be *related* to the background. Thus, it is an independence without isolation. When there is a focusing to the exclusion of other objects, and when the object is viewed from a narrow perspective, the object is isolated from the background. For example, when *S* concentrates on the ruler to the exclusion of other objects and the background, size constancy of the objects and veridical perception of the background is not achieved. When this occurs, the total field is not taken into account; i.e., the distance and surroundings of the objects are not related to the object. In contrast, at advanced levels of functioning there is a relationship of reciprocality; the perception of the objective distance and shape of the back wall is taken into account and influences the development of constancy of the object, and coordinately, the object develops constancy and thereby influences the distance and shape of the back wall. (Cf. Schlesinger's conceptions [17] concerning focusing vs. nonfocusing ability in terms of cognitive-personality functioning.)

15:20 (The ruler is moved L and R.) The window widened and lengthened.

16:25 (The ruler is moved L and R.) The right window grew larger, the left window smaller, but there wasn't much difference in the change, as before. I'm seeing them better. Before, the right window really widened out, but I was seeing it pretty good now.

With increasing development of the perceived trapezoidalness of the window frames, the window frames become stabilized, and the influence of the ruler on the window frames becomes less.

18:40 (The ruler is moved L and R.) The ruler stays about the same size. The left window is quite a lot smaller. When it went out, it got a little smaller. The second time I compared it with the ruler, and the windows got larger and flared out. I took them both in together. There was a comparison. The first time I just looked at the ruler.

The role of simultaneous comparison of object to background in facilitating veridical perception is noted.

d. Stick-pointing demonstration. In this demonstration a seven-foot bamboo pole is given to *S* with the instruction to touch various parts of the back wall with the end of the pole. For example, "Touch the upper left corner of the back wall and then swing the stick over to the right and touch the upper right corner of the back wall, etc."

At first a lack of correspondence between the way the room looks to *S* and the way it really is leads to maladaptive motor behavior and lack of motor coordination by *S*. The *S* either overestimates or underestimates the distance of the various target points on the back wall. With practice, however, *S* improves in her motor performance and forms new behavioral habits, but "the stick doesn't help me in seeing it." While the development of new percepts influences S's motor performance, there is not much direct influence of the motor system on the perceptual system. Even though S's motor behavior leads to an intellectual awareness of size and distance, her motor behavior alone does not lead to a perceptual resolution of contradictions. For example, by pointing to the four corners of the left and right windows in succession, S could kinesthetically feel a size difference and could thus become intellectually aware of the difference between the two windows. However, this awareness does not directly influence how the window frames look.

The effect of motor behavior with S-2 is that it appears to confirm or verify what S can actually see and also helps S to become more cognitively aware of what the nature of the distortion may be. It thus plays an important role in helping S to become more sensitive to new cues and increases the S's "openness" to new perceptual information. In this manner new cues may more readily be integrated into the perceptual system and new percepts may more rapidly develop.

We should note that with S-1, her motor behavior appeared to influence her perceptions to a somewhat greater degree, although this influence appeared to be quite unstable. However, it is difficult to speak of the effects of motor behavior in isolation, since visual cues also are presented and made available to S when she performs motorically.

e. Man and chair demonstration. In this demonstration *E* walks into the space between D and D′ and walks next to the back wall from D to C and then from C to D (see Figs. 3 and 4). In another part of the demonstration a chair is placed in the space DD′C with the back of the chair closest to C. In still another part *E* sits in the chair when it is placed in the space.

The *S* initially perceives the back wall in the plane C′D′; i.e., the frontal plane. When a person is placed in the space DD′C, the back wall cannot be in the plane C′D′. The back wall then must be behind the object, localized somewhere in the space DD′C. This is what occurred with *S*-2. Thus, a person in the space provides a strong overlay cue; i.e., when one object is perceived in front of another object, or in this case in front of the background, the background must be farther away than the object itself. However, even though *S* becomes aware of more space, the space is at first diffuse and the exact localization of the back wall is not well-defined. Thus, when *E* initially walks from point D to C, and therefore closer to *S*, a changing retinal size is still perceived as a changing objective size, and not as a constant size with a changing distance. The awareness of distance is undifferentiated.

However, with increased experience with the overlay cues, the perception of greater distance to the back wall becomes better defined and this has an effect upon the perception of the windows; i.e., the windows are perceived as farther away with a consequent change in size. Without the support of the overlay cues (when the chair and *E* are removed from the space), the perception of distance is decreased and the wall is perceived as closer. As the perception of space becomes more differentiated it is perceived as more normal and it gradually develops more stability and definiteness. At advanced levels of development there is not a sudden shift in perception without cue-support. Also, the space is perceived as an empty space with boundaries and with particular functional dimensions; i.e., as a space that can accommodate an object of a particular size.

Also noted during the development of the perception of the space is the conflict between cues. There is a conflict between accommodation for greater distance brought about by the overlay cues and accommodation for a lesser distance brought about by the initial localization of the back wall in the frontal plane. The nonresolution of the conflicting cues is manifested in the perceived blurring and fuzziness of the back wall and of the window frames.

f. Rectangular card demonstration. In this demonstration a large rectangular-shaped card made out of poster board is placed by *E* in various positions in the cubical section and along the back wall.

When *S* compares the card with the standard provided by the *E*'s hands, the card maintains its shape constancy; i.e., it is perceived as a rectangle at an angle. Also, when the card serves as a standard for comparison with the window frames, the window frames themselves begin to change toward their objectively trapezoidal shape. Without the comparison-standard of the card, however, the window frames change back to their previous rectangular shape.

However, a little while afterwards in the experiment, some kind of primitivation process seems to have been taking place. The rectangular card, when held in various positions and even in the frontal plane, is then perceived as trapezoidal. Only when the card is related to a stable standard (e.g., the vertical or horizontal surfaces of the cubical section), is veridical perception of the card obtained. It suggests that a regression in perceptual functioning had taken place, since previously the card had maintained enough stability to serve as a comparison-standard for the window frames. This form of instability of a percept and apparent reversal of a developmental trend is not uncommon in our protocols. It is usually an indication of a transitional stage between an old and a new form of perceptual functioning. In *S*'s attempt at a new form of functioning there is much groping and experimenting which gives the appearance of a regression, but which presages and prepares *S* for higher-level functioning. In the present case, it is noteworthy that the regression occurs during and immediately following the "storm" of the *S*'s development; namely, the period where *S* complains of tension, eyestrain, and blurring of the back wall (cf. the section on the total structure of the room). This appears to be a "high-water mark" or turning point after which the storm subsides and further development takes place at a rapid rate. This is the stage where earlier gains are consolidated to prepare the individual for further integration and development.

As the experiment progresses the card loses its constancy only when it is held close to the back wall. When the card is unstable and does not possess much potency as a standard, the card is distorted through the influence of the framework of the window frames. The framework of the window frames is more stable than the standard provided by the rectangular card. Therefore when the card is placed against the window frames, the window frames remain rectangular and the card itself is perceived as trapezoidal. The card is distorted to fit in with the perceived rectangularity of the window frames.

Different aspects of the framework are noted to have different degrees of potency. When the card is held flat against the window frames, the right window frame, being large and encompassing a larger portion of the card, "swallows it up" and is a more potent framework than the smaller window frame. Consequently, the right window frame maintains greater resistance to change and coordinately more readily distorts the card which it surrounds. In addition, instead of the card's always being distorted to a trapezoidal shape, the *shape* of the card at times maintains its constancy, but at the expense of the perceived *angle of orientation* of the card; i.e., both the card and the window frame are perceived as rectangular, but the orientation of the card is distorted. When the card is held parallel to the window frame, *S* states, "The card looks like it is slanted into the window, like it is tilted away from me at a greater angle. That is why it looks like they are not parallel. I make up the difference in that it is tilted away more." However, when the card is held away from the window frame and is not in close spatial proximity to it, the window frame loses its influence on the perception of the card. Decreases in proximity help to release the card from the influence of the surrounding framework.

Gradually, the card is perceived more independently of the framework and therefore achieves more stability. The nearness or farness of the card from the framework becomes a less important determinant of the perceptual outcome and plays a less potent role in the figural effects. This is in contrast to early levels of development where proximity was a potent force in determining the degree of influence that the surrounding framework had on its enclosed objects.

At advanced levels of development the card develops more potency and becomes a more stable force in serving as a standard for comparison with the surrounding window frames. The card then maintains its constancy and is perceived as rectangular while the window frames change and are perceived as trapezoidal. When *S* relates the card to the window frames by simultaneously concentrating on and comparing the card with the window frame, veridicality of both card and window frame is obtained. Finally, when *S* perceives the card and window frames in a veridical and stable manner, there is less need for support from the comparison-standard in maintaining the perception of the trapezoidal windows.

g. Trapezoidal card demonstration. In this demonstration trapezoidal-shaped cards of different sizes, but of the same degree of trapezoidalness as the window frames, are placed in various positions inside the cubical section and along the back wall by *E*.

When the card is not perceived independently of the framework of the window frames ("when it becomes part of the window"), the card itself loses its trapezoidal shape and is distorted by the framework of the window frames. The card is then perceived as rectangular. But when the trapezoidal card is utilized as an anchoring point and is compared with the window frames, the window frames then are perceived as trapezoidal. At this level of development the perceived trapezoidalness of the window frames requires the cue-support of the standard provided by the trapezoidal card.

With the trapezoidal card there also is an indication of a primitivation of the percept, with the trapezoidal card being perceived as rectangular not only at an angle of orientation, but even in the frontal plane. Gradually, however, there is a development toward greater veridicality, with the card maintaining its trapezoidal shape in all positions. It is through comparison with other parts of the room ("I widen my range of vision so that the thing becomes small, so I can get the shape of it in relation to you and the rest of the room") that veridical perception of both card and window frames is achieved.

At advanced levels of development, the window frames are perceived as the same degree of trapezoidalness as the card. However, size constancy is not complete; the card, as well as the window frames, are perceived as narrower when viewed at an angle of orientation.

h. Back wall movement demonstration. In this demonstration the back wall is moved in an arc with the pivot point at CC′ so that the angle of orientation of the wall is increased or decreased. Movement of the back wall permits *S* the opportunity to obtain different perspectives or views of the back wall.

Through the *S*'s taking into account and integrating the different perspectives of the back wall, the effect of movement of the wall is to facilitate perceptual change in the shape of the window frames. However, after the perception of the trapezoidalness of the window frames is stabilized, movement of the back wall does not influence changes in their shape; their shape remains constant and the distance changes.

i. Binocular vision demonstration. In this demonstration *S* removes the eye patch from her left eye and looks into the room with both eyes.

When *S* views the room with two eyes, she does not immediately perceive the veridical nature of the room or the objects. When *S* looks at the whole room, the configuration of the square room maintains its stability. Only when the total configuration is perceived in its parts does the perception of the distortion of the room and of the veridical nature of the objects develop. Thus, even with binocular vision a process of development takes place.

As the perception of the distortion of the room becomes stabilized, the utilization of part-perceptions in facilitating the development of a veridical percept is not required. At this stage of development the percept is well integrated.

At advanced levels of development there is a decrease in the difference of perceptual outcomes with binocular and monocular viewing. The *S*'s perception of the room monocularly is about the same as with binocular viewing. However, less straining and effort seem to be required to maintain a given level of achievement with binocular cues. In addition, binocular viewing, being a potent cue in the hierarchy of cues, appears to add the phenomenal feeling of "reality" to the perception of the distortion. Cf. Michotte (13) in his discussion of the phenomenology of reality.

j. Outside light demonstration. In this demonstration a light in a reflector is moved around outside the room in such a way that its beam is directed through the space on to the back wall. Because the beam of light is partially cut off by the ceiling, floor, and right side wall of the cubical section adjacent to the space, the light throws a combination of bright areas and shadows on the back wall.

Two cues for the distance and space are made available to *S;* first, shadows on the back wall provide cues for distance, and second, *S* perceives the light *in* the space, and therefore, the added light takes on "space-filling" qualities.

The influence of the light and shadow cues on the perceived structure of the window frames is at first negligible, since at early levels of development the space is perceived as indefinite and not well articulated. Even as the perceived orientation of the back wall develops, the rectangularity of the window frames presents a very stable configuration which resists changes. As the space becomes better articulated, the windows very gradually change in size and shape. With the development of veridicality the perceived structure of the windows attains stability and requires less cue-support to maintain a given level of achievement. At advanced levels of development the percept remains stable and invariant despite variations in shadow cues.

k. Fluorescent light demonstration. This demonstration is similar to the outside light demonstration, in that a light outside the room throws light and shadows on to the wall. However, in the fluorescent light demonstration, the light directed to the back wall is white, and thus, in addition to changes in brightness and shadow values, there are changes in the color values of the back wall. The results are similar to those in the previous demonstration.

l. Light variation demonstration. The standard illumination for the room is provided by (*a*) a light source suspended in the cubical section in a special ground-glass diffusing mechanism, whereby proportionately more light is directed to the distant part of the back wall than to the near part (to make up for the increased distance to the far part of the wall) and (*b*) indirect lighting directed through the space on to the back wall. Both are controlled by a variac. In the light variation demonstration the illumination is increased, decreased, or varied up and down consecutively by varying the standard setting on the variac. Therefore, increases or decreases in illumination do not change the brightness relationships on different parts of the back wall, but just change the total brightness value; i.e., the back wall gets either brighter or dimmer than standard illumination.

The *S* at first perceives the back wall as "flapping back and forth"; i.e., as going farther away and nearer. An increase in illumination is perceived as a cue to distance (and then the wall flaps farther away), while a decrease in illumination brings the wall nearer. This apparent contradiction to the traditional brightness-distance relationship is resolved when we understand how *S* actually perceives the increases and decreases in illumination. An increase in illumination not only makes the back wall brighter, but fills up the space and the wall "moves back a little to let the light in." However, at times the "light-filling space" cues are not potent enough and the room remains square, while the variation in illumination is perceived only as a variation in brightness.

With increasing development the perceived space influences the perception of parts of the wall in the direction of veridicality. When *S* relates the window to the perceived orientation of the back wall (i.e., perceives the size and shape of the window in relation to their angle of orientation), the development of veridicality is facilitated. At advanced levels of development the room is perceived as distorted and an increase in illumination is perceived as an increase in brightness. Thus as veridicality and stability are achieved, the perceived structure of the room remains constant despite variations in environmental stimulation.

B. Transfer Experiment

In this experiment Room 1 distorted section is rolled away and Room 2 distorted section is moved into position in front of the cubical section which remains identical. While Room 1 is distorted out toward the right; i.e., the large gap is on the right side and the windows are trapezoidal with the widest side on the right, Room 2 is distorted out toward the left (compare Figs. 3, 4, and 5, 6).

When we come to the transfer experiment, *S* has experienced the long training program provided with Room 1 and has developed a new mode of perceiving Room 1. At the start of the training experiment, *S*-2, for example, perceived Room 1 as a cubical room; throughout the training there was a long and gradual development toward veridicality; and finally at the end of training *S* perceived Room 1 as it objectively was. In other words, *S* in the transfer experiment is not the fresh and naive *S* who has never previously experienced a distorted room. The *S* in the transfer experiment has built up a new perceptual and motor adaptation to the first "experimental world" we provided him. The questions we now raise are: What has *S* learned in Room 1? When *S* is confronted with Room 2, what will carry over from Room 1? How will *S* adapt to this second "experimental world" we provide for him; i.e., Room 2?

1. Posttraining series in Room 1. Since five months had elapsed between the end of the training experiment and the beginning of the transfer experiment for *S*-2, she was placed before the training room again (Room 1) to see how much of the previous learning had been retained. As was expected, *S*'s first perceptions of Room 1 after five months were not like her veridical perceptions at the end of the training experiment.

There was a compromise effect; the total structure of the room was perceived neither in terms of veridicality (i.e., distorted), nor in terms of the proximal projection (i.e., square). Moreover, when the objects or cue-demonstrations were placed in the room, they were not perceived as they naively had been perceived at first, i.e., as distorted, nor were they perceived as they were at the end of the training experiment; i.e., as constant. A regression from veridicality had taken place; the objects were not adequately related to the framework of the room. Therefore, several hours of further practice in viewing the various objects in Room 1 was undertaken in order to help *S* develop veridicality again. However, during this posttraining series the perceptual achievements of *S* were unstable and there was considerable fluctuation in the perceptual outcomes at various times. Concomitantly, motivational difficulties were apparent and it was consequently considered expedient to attempt to investigate the nature of these difficulties. For several hours *S* freely expressed her feelings about the experiment. The *S*, having previously worked quite hard to perceive the room as it was objectively, and having achieved this goal of being "in contact with reality," now presented some considerable resistance to the experiment. Her drive to perceive reality and her curiosity to see the unknown had been satiated. The *S* became frustrated with the repetition, and stated that she had reached the limits of her ability to learn in the present situation. She said that if there were only something new in the experiment, some new problem or puzzle which she could attempt to solve, she would be able to start out fresh and continue to explore the experiment with *E*. Consequently, because of these motivational considerations, even though *S* had not fully redeveloped a stabilized veridical percept of Room 1, the posttraining series was discontinued. Between sessions *E* placed Room 2, the transfer room, in position and thus provided a "new problem" for *S* to solve. Thus, unknown to *S*, the next session she looked into Room 2, just as she had on previous occasions looked into Room 1.

2. The transfer room. At the beginning of the session the cigarette package rests in the center of the upper string in the cubical section, while the back wall is completely bare. Not only does *S* perceive the room as square when the room is objectively slanted out to the left (at times there is some ambiguity and *S* even perceives the room in the direction of Room 1; i.e., slanted out to the right), but she also misinterprets the cue "built-in" to the cigarette package cord. In terms of perspective relationships, the cord on the back wall gets narrower toward the left side as it moves into the distance along the back wall. Not only does the possibility of utilizing the narrowness of the cord as a cue to the distance and slant of the back wall remain unrealized, but even though the cue is noted by *S*, it is misinterpreted, and the wall is perceived as being "farther away on the right-hand side."

E: All right; close your eyes a minute while I set up the next demonstration. (Long pause.) (The lower package is placed on the left side of the back wall.) All right, open your eyes. (Pause.) (The lower package is moved from L to R.)

S: Hey, there's something wrong! (The lower package is moved from R to L.) Hmm, I've been had. The wall's tilted the other way!

E: How can you tell?

S: Well, I can tell the way the cigarette package moves. Move it, see if it will go on beyond the wall.

E: Well, let's look at the other cigarette package. (The upper package is moved L and R inside the cubical section.)

S: Well, that's just as it always was, I think.

E: (The lower package is moved L and R.)

S: Umm, I can't figure it out. The wall doesn't look tilted that way. You can't bend the thing in the middle can you? It can't be tilted away though on the right, because I saw the cigarette package. (Pause.) Hmmm. (Long pause.)

E: All right, I'll move the cigarette packages again. (Upper and lower packages are moved L and R.)

S: It has to be tilted away. I'm sure it has to be because that cigarette package looks just the opposite as it did before when it slanted downhill to the right—now it does the same thing to the left. (Short pause.) And it looked better to me today than it did yesterday.

E: Better? How?

S: Well, I've seen space over on the right—I thought I was. (Short pause.) I give up.

E: You give up. Why?

S: (Laughter) I'm completely confused now. The way I can see it, it looks better today—apparently just backwards.

E: Um-hmm. Let me move the cigarette packages again. (The upper and lower packages are moved L and R.)

S: I don't see anything, except that they're much farther apart down at the end. Ummmm—(Pause).

The perception of the wall in the frontal plane (or slightly slanted out to the right) and the perception of the cigarette package moving into the distance toward the left immediately presents a conflict to *S*. However, the conflict is not resolved in the direction of distorting the perception of the package, as in the initial perceptual outcome in the training experiment. On the contrary, the package is perceived as it objectively is; i.e., moving downhill into the distance. That the conflict is not completely resolved, however, is reflected in the lack of relating the perceived qualities of the package with the perceived nature of the back wall. Even though the package appears to be moving out toward the left, and therefore the wall "has to be tilted away toward the left," the wall does not immediately look tilted that way. The *S* still believes the wall to be the Room 1 wall, since the window frames appear rectangular or slightly trapezoidal to the right. Only when other cues for the perception of the window frames are subsequently presented does *S* integrate the perceived orientation of the wall with the new shape of the window frames. It is this lack of integration which presently provides for the feelings of frustration and confusion. Also interesting is the fact that the objective nature of the moving package is perceived "better today"; i.e., *S* not only carries over the operation of perceiving the moving package as moving-into-the-distance, but perceives it even better in a new situation, where the satiation effect has not yet set in. This suggests that there may be important relationships between the nature of the perceptual outcome and the motivational setting in which is occurs.[5]

E: All right. Close your eyes a minute while I set up the next demonstration. (The designed cards are placed along the tops of the window frames.) Open your eyes.

S: Well, I guess it's tilted away to the left. (Pause.)

E: How do you guess?

S: Umm, the cards look as if they're—well how do they stand up? The thing looks like it's reversed. I don't see how you can do that unless you have two walls. Yeah, it is! That right window is smaller! (Pause.) No wonder I couldn't see a peak up in the corner.

E: A peak, what do you mean?

S: I looked for that peak up in the right hand corner and I didn't see it. But I didn't see that there was one in the left hand corner. (Pause.) I couldn't see that at all when I looked in. It's very plain now that the windows are reversed—that the small one is on the right. They even look like trapezoids.

E: They look like trapezoids.

S: Um-hum.

E: Can you tell me something—how you first saw them as trapezoids?

S: When you put the cards up.

E: Yeah, what happened?

S: Well, then I saw that they slanted up and away to the left. And then I saw that—well I really took a good look at them then I guess—maybe then it dawned on me that the left one was larger and that the wall is really reversed, you know, completely.

E: Before I put the cards up you thought that the wall was tilted out on the left but it was still the same wall?

S: Um-hum, yeah, that's what I thought. (Pause.) And I was sure those two cigarette cords were farther apart over there on the right, and they're not at all.

When the possibility of there being two walls is recognized and accepted by *S*, the development toward veridical perception proceeds at a rapid rate. The addition of the cards to the tops of the window frames very quickly facilitates veridical perception of the room; i.e., the wall is perceived as tilted out toward the left and the windows are perceived as trapezoidal toward the left, the left one being larger than the right. As with the cigarette packages, the cards are not perceived as distorted to fit in with the perceived rectangularity of the window frames, as in the beginning of the training experiment. In striking contrast, in the transfer experiment the cards maintain their objectivity and serve as a standard for and are integrated with the perception of the window frames. The advanced level perceptual functioning acquired in Room 1 is utilized and adaptively generalized to the new distorted room. The development toward veridicality is attained at a rapid rate in the transfer room and much earlier than in the training experiment. Comparable levels of achievement were reached within one hour in the transfer experiment as compared to approximately 20 hours in the training experiment.

E: All right. Close your eyes. (The cards are taped below the edges of the bottoms of the window frames.) (Pause.) Open your eyes. What do you see now?

S: Two rows of cards instead of one. Slanting away to the left. (Short pause.) That's funny, I can see it better—or maybe it's the cards that make it easier to see. But I can see it better this way than I did the other way now.

E: Um-hum. In what way can you see it better?

S: Well yesterday—or whenever we started this thing, I don't remember how many days it was—but when you put those cards on the bottom it didn't do a thing to those windows —nothing at all. But it does when you do it this time.

---

[5] This is not the place to discuss the role of motivation in perceptual learning, but the following may be noted. Under high motivation *S* appears to become less distractible and becomes sensitized to the cues and thus becomes better able to utilize a given cue-organization. He becomes more goal-directed. When an *S* undergoes strain, he probably is attempting to keep out the distracting aspects of the situation. There is a conflict of cues, and *S* has to resolve the conflict by relegating certain cues (perspective cues) to the background by giving less potency to them. When he exerts more effort, he attempts to give certain cues (object cues) more potency.

E: Um-hum.

S: Not so much for the left one as it does for the right one. The right one, I can see the slant of that one very clearly, even down at the bottom. I mean I think I can—that's the way it looks to me, but the room looked one way to me when I first looked in today and now it's another way, so I'm kind of not sure. Anyway that's the way it looks. (Short pause.)

E: All right, close your eyes again while I set up the next demonstration. (The cards are taped to the left side of the left window and the right side of the right window.) All right, open your eyes again. (Short pause.)

S: Some more cards. Yeah, these on this side are very much closer to me than the others—I mean the ones on the right side are closer than the ones on the left.

Similarly, the addition of the cards on the bottoms and sides of the window frames is readily integrated with the almost completely developed veridical perception of the room.

During the remaining parts of the transfer experiment, the other cue-demonstrations were presented. For example, when *E* walked into the space, there was no perceived strangeness of the person, nor was there any reported change in the appearance of the room. This provides another contrast to the initial perceptions in the training experiment (see man and chair demonstration) where the perceived size of the person changed and where the overlay cue provided for the increased awareness of the space and distance. This part of the transfer experiment is, thus, quite similar to the highly developed aspects of the training experiment. A degree of stability and veridicality has been reached in the transfer experiment in which the perceived structure of the room remains fairly constant despite variations in the environmental stimulation, denoting the rapid utilization of advanced levels of perceptual functioning.

E: All right, here's the pointer. Try to point to the upper right-hand corner of the room. (Pause.) Now the upper left-hand corner of the room. How did it feel when you reached the point?

S: Well, it was a little bit— I underestimated it a little bit on both corners, but not—uh, I didn't feel lost, but I just didn't quite see it accurately.

E: O.K. Try tracing out the window panes with the pointer.

S: I'm just trying to get my arm out. (Short pause.) I can't (laughter).

E: You can't what? (Short pause.)

S: I can't coordinate what I just saw with what I've been doing. I'm awkward.

E: What seems to be the trouble?

S: Well, I learned to follow that other outline so well I could do it without even looking at it, you know, and uh, this, even though I saw it turned away, when I take hold of this pole I still want to trace the old pattern, you know. I may have seen something, but it hasn't communicated itself yet. I don't think that makes very good sense. This is kind of left-handed work or something—it's kind of backward or something. Well, that's funny. I can see it that way but I can't uh—maybe I'll learn to—I don't know.

E: (The *S* practices pointing at various parts of the back wall.) How does it feel now that you've had an opportunity to practice with the stick?

S: Well, it isn't so bad. I haven't lost myself, but I still want to go the other way, you know just out of habit.

The *S*'s initial motor behavior in the transfer experiment is not well-coordinated with the percepts. However, it is not that *S* is "lost," as in the beginning of the training experiment, but that *S* carries over the specific motor habits from the "old pattern" which are ill-suited to the presented differently-shaped transfer room and which lead to maladaptive motor behavior. The *S* notes the strangeness of the situation in which she "can see it that way" (i.e., tilted out to the left), but this veridical percept is not communicated to her motor adjustment. Thus, the carrying over or transfer of specific motor habits from the previous training at first leads to maladaptive consequences in a new situation, since the old behavior is no longer applicable to the new distorted room. The *S* has to learn new behavioral habits in the new distorted room.

When other cues were presented, veridicality was maintained: (*a*) When the ruler was passed along the back wall; (*b*) when the back wall was moved in an arc so that the angle of orientation was increased and decreased; (*c*) when shadow and illumination cues were presented; (*d*) when the cards were removed from the window frames; (*e*) when the cigarette package was removed from the back wall. Throughout all these changes in environmental stimulation, the room was perceived as it objectively was and remained invariant. In addition, *S* could not see the room as square with rectangular windows when she attempted to do so. These perceptual outcomes denote the utilization of levels of perceptual functioning similar to those utilized during the advanced stages of development in the training experiment.

However, while advanced-level perceptual functioning was adaptively utilized and very early resulted in the achievement of veridicality, the total consolidation process involved in newly-developed percepts may be a somewhat slower and more gradual achievement which comes about through increasing contact and familiarity with the stimulus material. For example, with *S*-2, while a considerable amount of stability had been attained early in the transfer experiment, this stability later broke down, and during later hours the perceptual outcomes were variable and veridical perception was not always attained. The situation was, however, again complicated by motivational considerations, since satiation effects similar to the effects in the post-training series were manifested. Thus, the rapid generalization process may be contrasted with the gradual consolidation process in which various relationships are integrated and stabilized.

V. DISCUSSION

The *S* looked into Room 1 at the beginning of the experiment and reported a square room with rectangular windows of equal size. In order to investigate how adults could learn to perceive the veridical nature of the distorted room, we provided *Ss* with the opportunity to view objects and people as they moved about in the room and also to manipulate a stick and touch various parts of the room with it. For example, one object, a cigarette package, was attached to the cord on the back wall and was moved along the length of the wall. Since the back wall was at a 24° angle of orientation to *S*, as the cigarette package moved along the wall it moved toward or away from *S*. This situation in the distorted room produced an impressive deviation from size constancy. As the cigarette package objectively moved toward or away from *S*, it was perceived as either growing or shrinking in size and moving across the frontal plane. The same object was perceived as changing in size but at a constant distance from *S*. It is not surprising that this occurred, since with the room perceived as square, the wall was perceived to be in the frontal plane. Distance was the invariant and size was perceived as varying with retinal angle. Since the cigarette package as it moved toward or away from *S* changed in retinal size, it was perceived as changing in object-size. When the package was at the far end of the back wall, the package subtended a retinal angle much smaller than the package in the near corner of the room, and it was therefore perceived as abnormally small. Differences in retinal angle were perceived as differences in object-size and not as differences in distance.

As the experiment progressed, more and more objects were introduced into the room. Other objects underwent distortion similar to the distortion of the cigarette package when it was perceived in the context of the square room. For example, the designed cards were distorted in shape and size; i.e., objectively they were rectangular with square designs, but they were perceived as crooked with diamond designs. Rectangular and trapezoidal poster cards placed in the room also were distorted in shape and size. A man walking into the room changed size as he moved along the back wall. A yardstick grew and shrank as it is moved along the back wall. A marble rolling down a trough along the back wall was perceived as rolling uphill, and so on. (This last demonstration was used only with *S*-3.) These distortions brought about conflict in *S*, i.e., a conflict between perceiving the distortion of the room and perceiving distortion of the objects. The conflict was initially resolved in the direction of distorting the objects in the room and, as a consequence, the usual constancies and causal relationships of everyday life broke down. When more objects were provided and when *S* performed inadequately in judging the distance of different parts of the room with the stick, more contradictions were presented, and therefore motivation to attempt to bring about some kind of resolution of the conflict and to bring harmony into the perception of objects and room was increased. This was especially evident when a human being walked into the space along the back wall. People possess a constancy that is important for our interpersonal relationships with them, and therefore it was quite difficult to intellectualize or rationalize the perceived size changes of the person as he moved about the room.

A. The Percept Began to Change

As the experiment progressed, changes in the retinal size of the object began to be perceived not as changes in the size of the object itself, but in terms of cues to changes in the distance of the object. There was a gradual trend toward perceiving the object as constant-in-size-at-a-given-distance. Similarly, changes in retinal shape began to be perceived not as changes in the shape of the object itself, but in terms of cues to changes in the position or orientation of the object. There was a gradual trend toward perceiving the object as constant in shape-at-a-given-orientation. In both cases there was a change in perceptual attribution, with the varying aspects of the retinal distribution being attributed to the spatial surroundings of the object instead of to the object itself. Instead of referring retinal variations to the object, *S* referred them to the distance and position of the object. Relating the object to its spatial surroundings and taking into account the object's distance and orientation helped to bring about veridical perception of both objects and room. Thus object-constancy was not developed in isolation. The object and the surroundings developed mutually; as object-character developed enough potency to serve as a cue to its position in space, the position in space affected the nature of the object-character. As object-cues were given increasing potency, the nature of the object became the invariant, and the spatial surroundings became the variant to which the object was related and coordinated. Finally, veridical perception of both object and room was achieved.

When we say that objects were perceived in terms of the perceived squareness of the room, this means two things: First, that *S* perceived the room in terms of the perspective relationships built into the room, and second, that the objects were perceptually embedded in the framework of the perspective relationships of the room, or else the objects would not be perceived as distorted. Therefore, to utilize object-cues, *S* had to differentiate retinal variations of the object from the perspective relationships built into the room. The *S* had to perceive the object independently of its surrounding framework, so that it could be adequately related to the framework and thus serve as a cue to the nature of the spatial context. The retinal variations could then be referred not to the object itself, but to the spatial context of the object. It was, therefore, an independence without isolation, since by referring the object to the spatial context, it could influence the perceived structure of the spatial context. When the object was perceived as a segregated entity, it was not related to, nor did it serve as an anchorage for the spatial surroundings.

The relationship of the object-cues to the spatial framework was, however, a reciprocal one. As the spatial framework itself became more articulated and better defined in terms of its objective nature, it in turn influenced the perception of the objects in the room. That is to say, the development of the spatial framework toward veridical perception facilitated the development of object constancy, since when an object could be localized in the spatial framework, it could then be perceived as constant-at-a-given-distance-and-orientation. As the object developed constancy it could reciprocally be increasingly utilized as a cue to the structure of the spatial framework, and then the developmental cycle began over again. It was through increasing differentiation of objects and framework and yet progressive integration and coordination of objects and framework that veridical perception was facilitated. These conceptions of development in terms of differentiation and integration may be compared with similar ones developed by Baldwin (4), Murphy and Hochberg (14), Piaget (15, 16), and Werner (20).

The problem of learning to utilize retinal variations of objects as cues to the nature of the room, and reciprocally, as cues to the nature of the object itself therefore also was a developmental process. In other words, the cues that helped *S* develop veridical perception of the room and objects were not "given" to *S*, with S being only a passive recipient who incorporated the cues into his percepts and somehow accommodated himself to the room. The *S* had to assimilate himself to the room by actively disclosing and discovering what the cue value of the object was.

The utilization of object-cues did not have to be a process that developed with full knowledge of the usual and familiar characteristics of the object. If an object moved in space, a completely unfamiliar object could be utilized as a cue to distance and position. For example, as an object moved along the back wall toward or away from *S*, retinal size continually changed with distance. The *S* needed only to perceive thing-identity for changes in retinal size to serve as a cue to changes in distance. That is to say, an object remains the same object and does not become enlarged or contracted as it moves about in space, and therefore changes in retinal size had to be due to changes in distance. (The question of how thing-identity emerges in nascent perception is a long and complicated process itself that takes us beyond present experimental considerations, but it could be assumed to be operative in adults in the present experiment.)

It was through coordinating and interrelating retinal variations of the object as it moved about that *S* could determine whether the object was small and near or whether the same object was actually large and farther away. This was probably the case in developing veridical perception of the rectangular and trapezoidal poster cards, since they were initially both unfamiliar as to shape and size. However, the utilization of retinal variations of the object as a cue to its distance and position could be facilitated when the objects were familiar ones. With objects of known dimensions, even the object was static, the individual could attribute the known size and shape to the object, and thus utilize known size and shape as cues to distance and position. When the cigarette package, for example, which was familiar in size and shape, appeared to be abnormally small, this could mean that the package was at a greater distance than it actually appeared. However, the process was not simple, nor was it only an intellectual process, since the familiar object still had to be perceptually related and coordinated to the spatial context of the room if it was to be perceptually utilized as a cue. But knowing how an object should look could sensitize *S* to the possibility of its utilization as a cue to the spatial context. It was in this sense that familiar objects facilitated the development of new object-framework relationships.

As *S* increasingly resolved the conflict of cues, the next problem he faced was the development of stability of the new perceptual achievement. Up to now *S* had been able to achieve veridicality by adequately relating the objects to their spatial surroundings through taking into account their distance and position. As the new percept was developing, *S* needed the support of object-cues for this relating and coordinating process to take place. They acted as a standard or anchor from which the spatial surroundings derived their stability. In the early stages of development the loss of the cue-supports could bring about the regression of an advanced-level perceptual achievement. Gradually, however, the newly-developed percepts became consolidated and stabilized. As different parts of the room became increasingly interrelated, they grew increasingly free from the need for cue-support. Finally, a stable level of achievement was reached so that the addition or subtraction of object-cues did not alter the perceptual outcome. The room came to be perceived in terms of its objective nature without any objects in the room. As *S*-2 stated, after 22 hours of experience she could not even see the room as square with rectangular windows when she had made an attempt to do so.

It appears that *S* developed and perceptually constructed a perceptual stereotype. The *S* developed a particular perceptual response that was elicited when the given stimulus distribution was produced on the retina. Before *S* had the opportunity to obtain various experiences in the room, the given proximal stimulus distribution had given rise to the perception of a square room. In striking contrast, after experiences in the room, the same proximal stimulus distribution gave rise to the perception of a distorted room, a room with a distorted back wall and trapezoidal windows. The stimuli available to *S* were precisely the same; the perceptual outcome was very different. The percept ceased to be uncoordinated with the distal stimulus, and achieved coordination with it.

Thus, the role of experience and contact with the environment was to make the individual more coordinated with the environment. A given proximal stimulus could give rise to a variety of percepts. Among all possible percepts that could be derived from the given proximal stimulus distribution, *S*, having experienced the particular kind of experiences in this particular distorted room, built up a new preferred mode of perceiving, i.e., a Room 1 preference. The percept was altered in terms of the qualitative nature of *S*'s interaction with the objects and room itself, i.e., the environment, and the nature in which his interactions had been integrated into his previous mode of perception.

This last point was further demonstrated with *S*-3. This *S* had learned to perceive the first distorted room as it objectively was. Advanced percepts became consolidated and perception of the distorted room was stabilized and no longer required cue-support. Immediately afterwards in the next session, without the knowledge of *S*, the differently-shaped distorted room was put into place. This second distorted room had a projection equivalent to that of the first distorted room. Room 2 was presented bare; no objects were present in the room. What happened was that Room 2 was perceived as if it were Room 1. That is to say, Room 2 was perceived as distorted, but distorted in the way Room 1 was distorted. Although this may at first seem surprising, it is understandable since the stimuli that came to *S* from Room 2 were identical with the stimuli from Room 1. In this case the newly-developed preferred percept was Room 1, and when the same stimulus distribution was produced on the retina, it was perceived as Room 1. Thus, the individual's initial interactions with Room 1 provided new experiences which led to a modification of the preferred mode of perception. A new perceptual stereotype was developed which was carried over to the second distorted room. These results corroborate Kilpatrick's experiment (10) utilizing the Ames distorted room, in which he demonstrated the development of a perceptual stereotype and its transfer to an equivalent distorted room.

The question arises: What were the adaptive consequences of this newly-developed perceptual stereotype? With Room 1 *S* perceived the room in terms of its objective nature, so that an equilibrium between the individual and the environment was achieved. However, when Room 2 was put into place, *S* perceived Room 2 as Room 1 and *S* then experienced new conflicts and contradictions. Thus, a new disequilibrium was established. It appears, therefore, that the perceptual stereotype developed in Room 1 was adaptive in the specific situation in which it was developed, and provided for equilibrium *only* in that situation. The newly-developed perceptual stereotype did not provide for an environment-adequate adaptation which remained invariant and stable despite transformations in the spatial context.

B. What Has Been Learned

The next phase of the experiment was to find what *S* had learned in terms of its generality and transposability. After many hours of experiencing Room 1, *S*-2 and *S*-3 had learned to perceive the room and the objects as they objectively were. In addition, the new percepts developed stability and the room was perceived in terms of its objective nature without the support of object-cues. However, several months had elapsed before the transfer experiment with Room 2 was begun with *S*-2, and therefore the perceptual stereotype for Room 1 which this *S* had previously developed had regressed. Therefore, when the transfer experiment was begun, this *S* did not manifest a well-defined and stabilized perceptual preference for Room 1. Consequently, when *S*-2 looked into Room 2 without knowledge that it was a different room, it was perceived as mostly square, but at times somewhat like Room 1. However, in the transfer experiment we were not only concerned with the nature of the manifestation of the perceptual stereotype acquired in Room 1. We were also concerned with the nature of the perceptual functioning that *S*s had acquired in Room 1, over and above the acquisition of a perceptual stereotype.

What was striking was that when objects were presented in Room 2, both *S*s maintained a constant relationship to object-character. When various objects were placed in Room 2, they were not perceived as distorted and in terms of the initial perception of Room 2 as they had been in Room 1 at the beginning of training. For example, at the beginning of the training experiment when the cigarette package was moved across Room 1, it was perceived as shrinking in size instead of receding into the distance. During the course of the training in Room 1, *S* developed object-constancy and the cigarette package was perceived as constant in size and receding into the distance, as it was objectively. In contrast, in Room 2 the cigarette package was not distorted as first it was in Room 1 but was immediately perceived as constant in size and receding into the distance, demonstrating in a striking way the transfer of advanced-level perceptual functioning that had been acquired during the long training in Room 1. Similarly, and also in striking contrast to their initial distortion in Room 1, the designed cards also maintained their constancy in the context of a new situation. Then very quickly the wall came to be perceived at an angle of orientation, and the windows as trapezoidal, and the total structure of Room 2 came to be perceived in terms of its objective nature. The object-cues were quickly coordinated and integrated with the spatial surroundings of Room 2. Thus, Room 2 was perceived in terms of the stable and enduring characteristics of the objects within it.

It appears, therefore, that in addition to a perceptual stereotype *S* had developed a *generalized way of perceiving object-framework relationships,* so that the objects could serve as invariant anchorings and could be perceived as invariant throughout transformation in the spatial context. The new way of organizing the potency of cues acquired in Room 1 was effectively generalized to Room 2 and provided for the achievement of veridicality in Room 2. The newly-learned operation of organizing cues in a certain way was not restricted to functioning in a specific situation and was not tied to specific structural contents. Thus, not only did the newly-developed way of perceiving acquired in Room 1 make for the achievement of veridicality in Room 1; it also was carried over and adaptively utilized in Room 2 and provided for the achievement of veridicality which survived the transformation in the spatial context. The newly-developed way of organizing object-framework relationships provided for a generalized environment-adequate equilibrium, since invariant objects were perceived as invariant; both Rooms 1 and 2 came to be perceived in terms of their objective natures.

These conceptions in terms of the organization of cues may be compared with those of other workers who have considered the process of perception. For example, Ames states that, "The process (which takes place below the level of awareness) by which the innumerable cues are integrated into a perception . . . is not a simple summation of cues. Some cues may not be made use of at all or may be only partially made use of. In this process of integration the use made of cues (or the weight given them) depends upon their prognostic reliability" (2).

Hilgard discusses "how patterns of experienced frequencies determine the interpretation of data present to the senses. A simple frequency theory of perception is like a simple frequency theory of habit formation. Both say that we do in the present what we most often have done in the past. Just as in a word-association test we reply to the word 'day' with the word 'night' because of the frequency of day-night association, so in perception we see the object that the presented cues have most often signified. . . . The main criticism (of a simple frequency interpretation) is that in perception, as in learning, interpretations are modified by the patterns of experienced frequencies and not solely by the totals of such frequencies" (9).

Another investigation dealing with the organization of cues is that of Weiner (18), where *Ss* learned to reorganize the potency of visual and body cues in the perception of the upright. The newly-learned organization of cues led to a change in perceiving the upright and was effectively generalized to a second orientation situation.

C. *Perceptual Learning and Unlearning*

Up to now we have been dealing mainly with the acquisition of perceptual stereotypes and new ways of organizing object-framework relations. This was not meant to imply, however, that this exhausted the qualitatively different kinds or forms of perceptual learning processes that were possible and that may have contributed to the achievement of veridicality. For example, much of the learning was *unlearning*. "Ignorance," as Mark Twain would say, "is not, not knowing, but knowing things that ain't so." In the training experiment, giving potency to perspective cues in the distorted room maintained the perception of the square room and provided for object-distortion. Only when perspective cues became poor and unreliable cues and were given less potency, could other cues, such as object-cues, be given more emphasis and veridical perception be achieved. As one *S* stated after nine hours' experience in the room, "I almost believe what I see. Before I didn't believe what I saw *because I've learned not to*. . . . All I can say is that's what it looks like now." In other words, concomitant with the learning to give greater weight to object-cues, *S* had to unlearn the usual potency given to perspective cues in everyday life. Similarly, in the transfer experiment *S* had to unlearn the perceptual stereotype built up for Room 1 before new learning could occur in Room 2. The material to be unlearned impeded the learning of the new material.

The data suggest that *S* developed a new "cognitive outlook" which significantly influenced and played a role in his perceptual functioning. For example, in the early stages of development in Room 1 *S* would attempt to cling to his initial perception of squareness by rationalizing or intellectualizing contradictions. As a consequence, much energy seemed to be devoted to this effort and *S* manifested resistance to change. However, when *S* could reckon with the hypothesis that the room could in some way be distorted, *S* could accept contradictions and conflicts as a problem to be solved, and not as something which had to be "held onto" and rationalized. For example, after *S* had attempted to hit the back wall with the stick and was not successful in the attempt, she stated that, "Something must be wrong with the room." If one could emotionally accept the fact that rooms could be of all shapes and, as one *S* explained, if one could take a kind of questioning and open or receptive attitude toward his own percepts, it became possible to accept new ways of integrating stimuli, and therefore, possible to put less stress on perspective cues as a valid cue to the objective nature of the room. Thus, the acquisition of a new cognitive outlook may have influenced the way in which cues were organized.

Perceptual development encompasses *more* than learning to reorganize the potency of cues. Perceptual learning includes an *"openness"* to new perceptual information, tolerance for exploring new possibilities, and becoming more accessible to new ways of integrating stimuli. This new cognitive outlook also played an important role in the *generality* of what had been learned; i.e., the degree to which the newly-learned object-framework relationships or cue-organizations could be generalized to a new situation. For example, when *S* was initially confronted with contradictions in the transfer room, i.e., Room 2, not only could the conflict be emotionally as well as intellectually accepted, but also exploration of what the conflict was could be undertaken. In other words, even though a particular cue-organization was a means to solving the conflict, the greater task, and one which had general adaptive value, was clarification and definition of what the conflict was. The task of learning a transposable solution to a conflict or problem was thus one of *enlarging awareness of what the conflict or problem really was* so that a particular cue-organization could be utilized in the resolution of the conflict. For example, with a new cognitive outlook toward distorted rooms in general, hypotheses concerning the differently-shaped rooms that might be possible could more freely be entertained. Because many different shapes of rooms were possible, perspective cues again might not be reliable and the possibility of giving object-cues increased potency was enhanced. It appears that *S* learned to enlarge hypotheses about the possibilities of differently-shaped rooms and the potency given different cues. It is this openness to new perceptual information and increased sensitivity which may have helped to facilitate the transfer of the previously acquired way of organizing the potency of cues. In other words, *S learned how to learn,* and this may have served to facilitate the generality of perceptual learning.

These phenomena remind one of certain processes in psychotherapy, where the goals are to help the patient outgrow or unlearn his neurotic tendencies of learning how *not* to learn and to replace these with healthy tendencies of learning how to learn which would have generalized value outside the therapy situation. We also are reminded of Harlow's experiments (7) in which he noted the general adaptive importance of learning how to learn in the solution of new problem-solving situations.

These experiments indicate that defining the nature of the proximal stimuli (the raw data) was necessary, but this was not sufficient for a complete understanding of the operation of the perceptual system. The nature of the dimensions of what was mediated to the individual, i.e., the percepts, was not solely determined by the mediation itself, i.e., the proximal stimuli. To be sure, there could be no perceptual outcome without the proximal stimuli; it was a necessary link between the environment and what the individual perceived. But it was the *organization of the mediational processes* that determined what a given perceptual outcome would be.

D. The Nature of Perceptual Organization

As we saw in the experiments, organization was a mode of functioning, an operation or process that changed and developed with the ongoing experiences of the individual. Different levels of development manifested different ways of integrating stimuli from the environment and *S* gave different potencies to the different cues at these different levels. Thus, different levels of development possessed different modes of organizing the proximal stimuli, i.e., different organizing principles. The nature of the organizing principle affected the nature of the perceptual outcome.

Therefore, organizational processes were developing operations, and not preformed percepts or structures, or fixed patterns of operations, or static processes that emerged without a history. Perceptual development was understood as reorganizations of proximal stimuli and included new and developing operations in the perceptual system. Perceptual development did not imply only a quantitative increase or change of percepts, but also a development or a reorganization of processes.

In the development of the individual's perceptions, interaction and experience with the environment modified and transformed organizational processes and percepts. Interaction with the environment implied that the individual did not passively accommodate himself to the actions of the physical energy distributions impinging on the retina. The environment acted on the existing developmental level of the organizational processes in the individual; the impacts of surrounding stimuli did not act in isolation, but acted by helping to modify the individual's action on the environment. Thus, in the construction of ever more complex and articulated percepts, we should not expect to find the role of experience and organizing activities separated. We should not expect to find "experience" without a process of experiencing carried out by an *actively organizing individual.* A given organizational process could be modified by new experiences, and experiencing could be modified by the existing organizational process. It was an interaction of mutual dependence. Interaction with the environment helped to make the individual more coordinated with the environment by affecting the organizational processes, and the nature of any further interaction between the individual and the environment itself was affected by the processes of organization. These conceptions may be compared with the more extensive formulations of Piaget (15, 16) in his discussion on assimilation and accommodation.

For example, the perceptual and motor interaction obtained in the training program in Room 1 provided for the development of new percepts which were more coordinated with the objective nature of the distorted room. But beyond the development of these new percepts, *S* developed new forms of perceptual operations; i.e., new object-framework relationships. These new organizing relationships in turn provided for the emergence of new percepts not only in the situation in which the organizing relationships were developed, but in a new situation (i.e., both in Rooms 1 and 2). Thus, interaction in the room brought about the development of new organizational processes, and reciprocally, the organizational processes affected the nature of the perceptual outcome. Organizational processes were affected by and affected the nature of the individual's interaction with his surroundings.

How organization theory is utilized in current perception research is demonstrated by Klein's (11) "cognitive attitudes" and Witkin's (22) "perceptual modes," whereby they mean an individual's preferred style or principle of organizing a complex of stimuli in terms of early-developed and widely-generalized needs of the personality structure. For example, in Witkin's research systematic variations in perception were noted both developmentally (i.e., differences in perception varying with age) and interindividually (i.e., differences in perception between individuals of the same age) along the field-dependent–field-independent perceptual dimension. It appears that at the field-dependent end of the dimension, there is a relative lack of differentiation between body awareness and visual field; whereas at the field-independent end of the dimension, the body awareness has been differentiated from the visual field, but it has been given undue weight in the final integration of the body and the visual field. With most *S*s there is a differentiation and yet an adequate integration of bodily and visual cues, so that neither extreme of the dimension is reached. Witkin's research suggests that a particular organizing principle develops from the total life experiences of an individual and forms the basis for a particular integration of bodily and visual cues.

Thus, it may be suggested that differences in perception are due to differences in the organizing processes in different people, either because the process has been arrested at a given level of development, or because certain functions in a given process-pattern are given undue weight. This is especially noticeable in the perception of people where an individual's perception of an interpersonal situation may be distorted or constricted because of the nature of the existing organizational processes as they have developed in the life experiences of the individual. In other words, one way of considering the perception of people is that a percept may be environment-inadequate because a certain kind of organizing principle in the individual leads to distortions or impoverishments of the stimuli that may be available to him.

A psychophysics of perception is necessary for an organizational theory, but it is not sufficient. The stimulus conditions do not define the total operation of the perceptual processes; they do not define the nature of the organism's response to or integration of a given stimulus complex. Thus, organization theory does not, on the one hand, deny the importance of stimulus conditions, such as the orderliness of the stimulus field, stimulus gradients, and figural properties (6, 19). Nor does organization theory, on the other hand, deny the importance of experiential conditions, such as attitudes, expectancies, and motives (2, 18). Organization theory is a developmental field theory whose processes are dependent on *both* the operation of stimulus factors and experiential factors. Cf. Attneave (3) for a discussion of these factors in relation to economical principles.

These considerations raise a methodological point concerning experiments dealing with veridical perception. For if we examine cases of veridical perception, i.e., cases in which the percept is coordinated with the environment, we find out only this fact: namely, that the percept *is* coordinated with the environment. We disclose the nature of the organism's achievement. *But how this environmental-adequacy is brought about, and on what processes it is based, can be disclosed by investigating inadequacies.* In the following discussion we shall attempt to develop some concepts put forth by Heider (8).

An analogy may help us here. If one looks through a transparent pane of glass, one finds out about the objects on the other side, but not about the glass itself. When the pane does not simply mediate what comes from the other side, that is, when it is not a pure medium, like air, then one finds out about the pane itself. For example, when one touches the glass with one's hand, then one can get information about the pane itself. This is also the case when it is possible to see it from the side or when there is some defect in the glass. In a similar manner, percepts act as a medium between the environment and the underlying perceptual processes. The more the percepts seem to mediate the environment, i.e., the purer the medium, the less one can find out about the properties of the perceptual system. Then one finds out about only the properties of the environment. Only when the percept is a poor medium, and when it is environment-inadequate, then we do not find out about the environment but about the properties of the perceptual system.

We may represent these schemata in the following way. A percept *(P)* has its sources partly in the processes of the perceptual system *(p)*, and partly in the environment *(E)*. When *P* and *E* are coordinated, one finds out about *E*. However, when *p* produces inadequate *P*, one can get information about the properties of *p* from the nature of the distortion if one has independent knowledge of *E*. Those features of *P* which are not coordinated with *E* must have their source in *p* itself. Thus, when *P* and *E* are not coordinated, one infers from the effect *(P) to the* underlying cause *(p)*. It is in this manner that inadequate perception may reveal the processes of the perceptual system.

In our investigation we attempted to follow these conceptualizations as a methodological guide for experimental design. Consequently, a stimulus condition was set up so that initially *S* would manifest environment-inadequate perceptions. In this way we were able to learn, first, something about the structure of the perceptual system, from the nature of the initial perceptual inadequacies and how *S* related to the visual environment under these conditions. Second, we were interested in *S* as a developing and growing organism, not as a static being fated to his own distortions. How environment-adequacy was brought about, and on what processes it was based, we could find out by investigating the ongoing processes of *S* in his attempt to construct a percept that would form a harmonious and conflict-free relationship with the environment. It is believed that future investigations directed toward the nature of the organizational processes at various levels of development may more adequately define the nature of the individual's response to his environment; i.e., the nature of his perceptions and his interactions with his surroundings.

VI. SUMMARY

The present study was concerned with how adults develop new visual percepts and new modes of perceptual functioning. For this purpose a room specially designed for this study was constructed in which *S* would initially manifest maladaptive perceptual and motor behavior. For example, the initial perceptual outcome when *S* viewed the bare experimental room was a cubical room, when in reality it was distorted. In addition, when the situation was made more lifelike by placing objects within the room, *S* was immediately confronted with a conflict. There was a conflict between perceiving the distortion of the room and perceiving the distortion of objects in the room. The conflict was initially resolved in the direction of perceptually distorting the objects.

In order to investigate how adults learn to adapt to the strange and unfamiliar "experimental world" of the distorted room (i.e., to perceive the room and objects as they are in reality), we provided *Ss* with the opportunity to have a variety of perceptual and motor experiences with the people and objects in the room. While viewing the objects from varying perspectives in relationship to the background of the room, the three *Ss* intensively studied in this experiment very gradually developed (after 25 to 40 hours) a new perceptual adaptation to the "experimental world" which we had provided them. A new constancy of objects was developed. For example, initially when a cigarette package was moved along the length of the distorted back wall, it was perceived as shrinking in size, instead of receding into the distance as it was in reality. After experience in the room *S* came to perceive the cigarette package as maintaining a constant size and moving away from him at a slanted orientation along the slanted back wall. Thus, the perceptual outcome was that various same-sized objects at different distances came to be perceived as same-sized objects at different distances, as they were in reality, and not as different-sized objects at equal distances, as they were perceived before the experiences in the experiment. The *Ss* learned to perceive the intrinsic properties of the objects and concomitantly, the intrinsic properties of the room. The distorted room came to be perceived as it was in reality, that is, as distorted. Veridicality was achieved for both objects and room.

The next problem was: What has *S* learned in terms of its generality and transposability? To answer this question, two of the three *Ss* were presented with a differently-shaped distorted room, but having a projection equivalent to that of the first distorted room. The same objects provided in the first room also were presented in the second room. Under these conditions *Ss* maintained a constant relationship to object-characteristics. They did not have to go through the slow step-by-step developmental process as in the first distorted room. The second room was quickly perceived in terms of the stable and enduring characteristics of the objects within it. Thus, *Ss* carried over the object-framework organizing principle developed in the first distorted room to the second distorted room. The perceptual functioning or operations acquired during the perceptual interactions in the first room provided for an adaptive orientation which survived a transformation in the spatial context; i.e., the newly-developed perceptual processes were adaptively utilized in a new situation.

Throughout the research an attempt was made to analyze the detailed phenomenological protocols not only in terms of the developing percepts but also in terms of the developing processes and new modes of perceptual functioning. The problem of object-constancy and its development was considered in relation to object-framework relationships. The implications of the results for an organizational theory of perception also were discussed.

REFERENCES

1. Ames, A., Jr. Nature and origin of perception: Preliminary laboratory manual for use with demonstrations disclosing phenomena which increase our understanding of the nature of perception. Hanover, N.H.: Institute for Associated Research, 1946-47 (mimeographed).
2. Ames, A., Jr. Nature and origin of perception: Literature dealing with the significance of the phenomena disclosed by the demonstrations. Hanover, N. H.: Institute for Associated Research, 1948 (mimeographed).
3. Attneave, F. Some informational aspects of visual perception. *Psychol. Rev.*, 1954, 61, 183-193.
4. Baldwin, A. L. Behavior and development in childhood. New York: Dryden, 1955.
5. Duncker, K. Über induzierte Bewegung. Ein Beitrag zur Theorie optisch wahrgenommener Bewegung. *Psychol. Forsch.*, 1929, 12, 180-259. Trans. and condensed by W. D. Ellis (Ed.), *A source book of Gestalt psychology*. New York: Humanities, 1950. Pp. 161-172.
6. Gibson, J. J. *The perception of the visual world*. Boston: Houghton Mifflin, 1950.
7. Harlow, H. F. The formation of learning sets. *Psychol. Rev.*, 1949, 56, 51-65.
8. Heider, F. Die Leistung des Wahrnehmungssystems. *Z. f. Psychol.*, 1930, 114, 371-394.
9. Hilgard, E. R. The role of learning in perception. In R. R. Blake & G. V. Ramsey (Eds.), *Perception: an approach to personality* New York: Ronald, 1951. Pp. 95-120.
10. Kilpatrick, F. P. Two processes in perceptual learning. *J. exp. Psychol.*, 1954, 47, 362-370.
11. Klein, G. S. The personal world through perception. In R. R. Blake & G. V. Ramsey (Eds.), *Perception: an approach to personality*. New York: Ronald, 1951. Pp. 322-355.
12. Koffka, F. *Principles of Gestalt psychology*. New York: Harcourt, Brace, 1935.
13. Michotte, A. L'énigme psychologique de la perspective dans le dessin linéaire. *Bull.*

Classe Let. Acad. royale Belgique, 1948, 34, 268-288.

14. MURPHY, G., & HOCHBERG, J. Perceptual development: some tentative hypotheses. *Psychol. Rev.,* 1951, 58, 332-349.
15. PIAGET, J. *The origins of intelligence in children.* New York: International Universities Press, 1952. (French Ed., 1937.)
16. PIAGET, J. *Introduction à l'épistémologie génétique.* Paris: Presses Universitaires de France, 1950.
17. SCHLESINGER, H. J. Cognitive attitudes in relation to susceptibility to interference. *J. Pers.,* 1954, 22, 354-374.
18. WEINER, M. Effects of training in space orientation on perception of the upright. *J. exp. Psychol.,* 1955, 48, 367-373.
19. WEINER, M. The effects of differently structured visual fields on the perception of verticality. *Amer. J. Psychol.,* 1955, 68, 291-293.
20. WERNER, H. *Comparative psychology of mental development.* (Rev. Ed.) Chicago: Follett, 1948.
21. WITKIN, H. A. Perception of body position and of the position of the visual field. *Psychol. Monogr.,* 1949, 63, No. 7 (Whole No. 302).
22. WITKIN, H. A., *et al. Personality through perception.* New York: Harper, 1954.

(Accepted for publication November 28, 1955)

Vol. 70, No. 17 Whole No. 424, 1956

Psychological Monographs: General and Applied

Patterns of Religious Attitude Shifts from College to Fourteen Years Later

ERLAND N. P. NELSON[1]

University of South Carolina

I. INTRODUCTION

ARE there discernible patterns in shifts of students' religious attitudes from college to 14 years later? Do college attitudes toward religion tend to persist more than attitudes toward other values? Is there a significant relation between religious attitudes indicated on paper-pencil tests in college and overt behavior 14 years later? Do regional differences in attitudes tend to follow a pattern? May we expect to find consistency in shifting of belief in Deity as compared with attitudes toward the church or toward Sunday observance?

For such questions a longitudinal approach seems appropriate, in which subjects are tested in college and later retested by the same instruments. This is a report of such a study involving four religious areas. Initial testing was done in colleges and universities in 1936, and retest was made of identical subjects in 1950.

II. PROCEDURE

The Initial Study, 1936

Subjects (*S*s) for the original study consisted of 3,749 students who attended 18 colleges and universities in 1936. These students were enrolled chiefly in liberal arts colleges of state universities and church-related colleges. Our *S*s, 1,724 males and 2,025 females, were distributed over the four undergraduate classes from freshman to senior. The institutions included: four state universities, three located in the Midwest and one in the South; six Lutheran colleges; two colleges affiliated with the Society of Friends; and single institutions connected with each of the following denominations: Methodist, Presbyterian, Seventh Day Adventist, and United Brethren. Of the twelve denominational colleges, two are east of Chicago, one is in the South, and the others are in the Midwest (8).

Scales for measurement of student attitudes toward religion were selected from the Thurstone series: Attitudes Toward God (A Reality), Attitudes Toward God (Influence on Conduct), Attitudes Toward the Church, and Sunday Observance (9). Each of these scales has been calibrated by the method of equal-appearing intervals (10). Validity is based mainly upon internal consistency, self-ratings, and certain behavioral indices of religious attitudes. Reliability for each scale is .90 or better—all of which is based upon groups additional to the present *S*s. For convenience in comparison, all scales were adjusted so that high scores indicate attitudes favorable toward

---

[1] Acknowledgment is made to all who have cooperated in this study, especially to Dr. D. A. Worcester, Dr. M. K. Walsh, Dr. L. L. Thurstone for original use of tests, Naida Nelson, and the University of South Carolina.

Bureau Ednl. Psy. Research
DAVID ING COLLEGE
Dated
Accs No.

the religious value under consideration.

Anonymity in marking attitude scales is sometimes thought to be a factor in honesty of response (4). However, Ash (1) found that on political, economic, and racial issues his *S*s were not biased either way by signing or not signing attitude questionnaires. Corey (3) likewise found anonymity unnecessary in attitude study.

In the present study, the following precautions were taken:

1. In both the 1936 and 1950 phases, subjects were encouraged to indicate attitudes frankly with or without signing.
2. In the 1936 study, it was made clear that the student's attitude could in no way affect his standing at his college since the unscored papers would be sent directly to the writer and held strictly confidential.
3. As an additional device, a supplementary test was added calling for attitudes toward the college or university then attended. Yet on this device to "test the test" the utmost freedom of response seems evident. In these institutions, signed expressions indicated attitudes no less antagonistic toward the college then attended than did those which were unsigned. If restraint in expression is felt anywhere, it should be expected when invited to indicate attitudes toward the institution then attended. Yet signing or nonsigning seemed unrelated to criticism of college or university (7).

The 1950 Follow-up

Subjects for the 1950 follow-up had been reduced in number by death, hospitalization, and by their failure to keep alumni offices informed of rapidly changing addresses. By the time our attitude booklets were mailed, accurate current addresses had been reduced to 1,200. Of these 1,200 who had presumably received the 1950 test booklets, a total of 893 completed the tests and returned them in time for our calculations. Ex-student responses came from every state in the Union, from South Africa, Liberia, most countries of South America, Asia, and Europe. Occupationally, the 1950 respondents included 65 ministers, 180 teachers, 198 business men and women, 37 physicians, and smaller numbers of authors of books, university professors, federal officials, enlisted men and officers, and one bartender. Of these 893 ex-students, 454 were females and 439 males.

A sampling question might be raised: are the 893 ex-students of 1950 representative of the 3,749 collegians of 1936? That the question is not crucial to the present study is seen from the fact that scores of the 893 ex-students are compared only with their own 1936 scores—not with the means for the larger group. All 1936 data have been recalculated on the basis of the identical 893 who responded in 1950. We shall present mean differences, 1936-50, and percentage of attitude shifts for these same 893 persons on the same tests after the 14 intervening years.

However, to evaluate the extent to which the 1950 group is representative of the larger 1936 population, two methods were employed:

1. Using one of the scales, the respondents of 1950 were found to differ only slightly from the means of the entire group when all were in college in 1936.
2. After nearly 800 booklets had been received in the fall of 1950, two mail "urgings" were sent out which brought 93 "stragglers" who responded after February 1, 1951. These were dated and marked for separate study. If these "stragglers" did not differ significantly from the early respondents, it might be assumed that the other 300 who failed to get our booklets or who neglected to return them would likewise not differ significantly. Comparisons were made between "stragglers" and early respondents in three ways; 1936 means, 1950 means, and 1936-50 changes. In none of the three ways did the "stragglers" differ significantly from the early respondents.

Thus, these two devices fail to indicate that selective factors could account for the test and retest results to be presented.

Stability of scale values, concepts, and *meaning of test items* over the 14 years intervening between test and retest deserve consideration. It should be recalled that this study does not propose to compare the 1936 test results with those of a new generation of college students, but to measure changes in identical persons. However, to evaluate the extent to which item meanings and consequent item values might have changed for a new generation of students, the writer made a brief study in 1954. The scale, Attitudes Toward the Church, and Thurstone's judging techniques were employed. Fifty-one judges were secured from advanced students in educational psychology and measurement classes. Tabulations indicate slight shifts in value for 22 per cent of the items, the trend being slightly higher for some antireligious items of the scale, but the degree of difference was not significant. On the scale, God, An Influence on Conduct, significant changes in scale values were found for less than 5 per cent of the items. (A 1952 study [2] using 60 volunteers from elementary psychology courses in California indicates somewhat greater changes in item values.) Even with a new generation of undergraduate students as judges, the change in item values is relatively small. With judges 14 years out of college, the changes in religious item meanings might be expected to be even less. Apparently, religious concepts possess marked stability.

III. ATTITUDES TOWARD THE CHURCH

Degree of Attitude Change After 14 Years

It will be recalled that this scale, for both test and retest, has been revised so that data may be compared conveniently with the other religious areas. Hence, high scores, 8 to 11.0, indicate attitudes strongly favorable to the church as expressed in such items as "I think the church is a divine institution, and it commands my highest loyalty and respect." Low scores in the area, .5 to 3.0, indicate strong antagonism, as typified in an item such as "I have nothing but contempt for the church." For convenience, the data from the 39,000 marked items have been summarized in Table 1. The degree of attitude change from 1936 to 1950 may be inferred from the difference between the means for this group of 887 *Ss*. Both test and retest means indicate attitudes favorable toward the church, 8.9 and 9.2 respectively. The change in means from 1936 to 1950, .30, is significant at the .01 level and points to attitudes more favorable toward the church after the 14 post-college years.

TABLE 1

ATTITUDES TOWARD THE CHURCH AMONG STUDENTS AT 16 COLLEGES AND UNIVERSITIES IN 1936, COMPARED WITH THOSE OF THE SAME SUBJECTS 14 YEARS LATER
($N = 887$)

1936		1950	
Score	*f*	Score	*f*
10	56	10	84
9	482	9	615
8	220	8	102
7	71	7	48
6	33	6	22
5	10	5	4
4	11	4	7
3	2	3	2
2	2	2	3
1		1	
Mean	8.90	Mean	9.20
SE	.04	*SE*	.034
Diff.			.30
p			.01
r			.38
Shifts		*N*	%
No shift		453	51
Toward church		308	35
Against church		126	14

Attitude Shifts

But since differences in means may fail to show the full extent of actual change (shifts *toward* the church may be offset by shifts *against*), we have counted the number of *S*s who in the 14 years did not shift by so much as one scale interval. It will be noted that 453, or more than 51 per cent of this group, remained in their original position (i.e., made an identical score) over the 14 years of added age, marked by war, marriage, and other personal and cultural changes. While one might expect some relation between paper-pencil tests in college and a later retest, that over half of the *S*s should remain in the same class interval for so many years is surprising. Apparently attitudes toward the church expressed in college do tend to persist. Judging by percentage of *S*s unchanged, attitudes toward the church showed significantly greater stability than any of the other attitude areas, even including attitudes toward God.

We found 126 *S*s, or 14 per cent, whose 1950 attitudes were less favorable toward the church than in 1936. In contrast, there were 308 persons or 35 per cent who had shifted toward the church by one or more intervals. This is in accord with the significant change in means toward prochurch attitudes already noted.

Still another indication of the 1936-1950 attitude persistence is seen in the Pearson product-moment *r*, .38, *SE* .025. While some shifts in attitudes have been shown, the correlation between test and retest with 14 years intervening suggests educational implications. What a student believes in college can hardly be ignored or considered as of mere transient importance.

Types of Institutions

That student attitudes toward the church differ among the various types of institutions has been shown in other studies (8). But do *changes* in attitudes over postcollege years also vary when we consider institutional grouping? For the sake of brevity, we classify the 16 institutions into 3 groups: the 4 state universities, the 6 Lutheran colleges, and a group of 6 other colleges affiliated with the following denominations: Methodist, Presbyterian, United Brethren, Adventist, and Friends.

As will be noted in Table 2, the type of institution showing greatest prochurch attitudes in 1936 was the Lutheran group, lowest mean scores were for the state universities, and the colleges of the "other" denominations were between these two. The differences between Lutheran and non-Lutheran denominational groups were not significant. Differences

TABLE 2

ATTITUDES TOWARD THE CHURCH, BY TYPE OF INSTITUTION

(N=887)

Types of Institution	*N*	1936		1950		Diff.	*p*	No Shift (%)	Shift Pro Ch. (%)	Shift Anti Ch (%)
		Mean	*SE*	Mean	*SE*					
4 state univers.	190	8.55	.12	8.96	.09	.41	.01	46	41	13
6 Luth.	357	9.05	.05	9.34	.04	.29	.01	54	34	12
6 other	340	8.96	.06	9.18	.05	.22	.01	51	17	32

between state universities and the other two types were both significant. By 1950, the ex-students from all three groups of institutions had increased their mean scores by significant differences. While all three groups had become more favorable toward the church, the ex-students from state universities had come further than those from the other groups, but had not equaled either.

The percentage of "standpatters" was greatest (54 per cent) among Lutheran ex-students and least (46 per cent) among state universities. Of ex-students from state universities, 41 per cent shifted in the direction of greater favor toward the church, compared with 34 per cent and 17 per cent among *S*s from the two denominational groupings.

Regional Differences

Significant regional differences in student attitudes were found in the 1936 phase of this study. Students enrolled in Southern colleges were more prochurch than those in the East and the Midwest. In spite of the wide dispersion of alumni and ex-students from all institutions (students from Midwestern institutions now residing in the South, New England, and the West), regional differences of 14 years earlier were still noticeable. Ex-students from Southern schools remained more prochurch than those from the North, but the magnitude of the differences had declined.

The Sex Variable

While in college in 1936, the usual pattern of sex differences was evident. On attitudes toward the church, the mean for women was .33 higher than for men. While women remained more prochurch than men 14 years later, the difference in means had decreased to .14. In college, the sex difference was significant at the 1 per cent level. Fourteen years later the difference in means proved significant only at the 7 per cent level.

The means for both males and females changed in the direction of attitudes more favorable toward the church by differences significant at .01 levels. More women than men were "standpatters" (remained in the same scale interval on both test and retest), but more men shifted toward stronger prochurch attitudes.

Attitudes Toward the Church and Overt Behavior

That attitudes tend to persist for at least 14 years has been shown by the small differences in means on test and retest, by the 51 per cent of "standpatters," and by the test-retest correlation of .38. But may we expect college attitude scores to be related to overt behavior 14 years later? If prochurch attitudes indicated in tests in college tend to be translated into action, we should expect positive correlations between college attitude scores and such postcollege activities as church attendance and financial support of the church.

That college attitudes toward the church are related to church attendance 14 years later seems evident. Using the question, "Do you attend church most of the time?," we find a biserial r of .258, *SE* .04, between 1950 church attendance and college attitudes in 1936. While not large, this correlation is significantly greater than zero.

If attitudes are related to overt behavior, we might expect a relationship between prochurch views in college and the percentage of ex-students who 14

Bureau Ednl. Psy. Research
DAVID HA AINING COLLEGE
Dated ....
Accs. No ....

years later tithe—give to the church 10 per cent or more of their income. We find that 76 per cent of the tithers reporting in 1950 had been above the mean in attitudes toward the church while in college. Of those tithing in 1950, 31 per cent had been in the highest interval of the distribution while in college in 1936, 25 per cent of the tithers had been in the next highest interval and 16 per cent had been in the third interval.

IV. ATTITUDES TOWARD SUNDAY[2] OBSERVANCE

Degree of Attitude Change

We have seen that these 887 ex-students tend to shift toward more prochurch attitudes 14 years after the initial college tests. We now turn to attitude changes in regard to Sunday observance. Will a similar pattern obtain—significant shifts toward the more religious end of the continuum?

Test data have been summarized in Table 3 with low scores, 1-3, indicating strong antagonism toward Sabbath observance. At the upper end of the scale, scores 6 to 9 indicate attitudes strongly favorable toward Sunday observance.

The degree of attitude change from 1936 to 1950 may be inferred from the differences in the 14-year means. Both 1936 and 1950 means are in the neutral areas, 5.92 and 5.25. While small, this difference, .67, is significant at less than the .01 level. In contrast with the mean change *toward* prochurch attitudes, the significant change here is in the *opposite* direction, less favorable toward Sunday observance.

TABLE 3

ATTITUDES TOWARD SUNDAY OBSERVANCE AMONG STUDENTS AT 16 COLLEGES AND UNIVERSITIES COMPARED WITH THOSE OF THE SUBJECTS 14 YEARS LATER

($N = 893$)

1936		1950	
Score	*f*	Score	*f*
10		10	
9		9	
8	84	8	49
7	172	7	91
6	191	6	130
5	162	5	169
4	170	4	231
3	110	3	218
2	1	2	5
1	3	1	
Mean	5.92	Mean	5.25
SE	.05	*SE*	.05
Diff.			.67
p			.01
r			.43
Shifts		*N*	%
No shift		257	29
Toward Sunday		174	19
Against Sunday		462	52

Attitude Shifts

But since means may fail to indicate actual change, shifts *for* observance *offset* by those against, we turn to the percentage of individual changes. Two hundred and fifty-seven *S*s, 29 per cent of the entire group, did not change their attitudes toward Sunday observance by so much as one scale point. One hundred and seventy-four, 19 per cent, shifted to attitudes more pro-Sunday, while 462, 52 per cent, shifted toward attitudes less favorable to Sunday observance. Thus while the difference between means, 1936 to 1950, is small, there is marked contrast between percentages of student shifts, 52 per cent toward less Sunday observance with 19 per cent toward more.

The contrast in attitude-shifts between that toward the church and against Sunday observance may be summarized as follows:

---

[2] Sunday and Sabbath were used interchangeably due to the fact that *S*s included Seventh Day Adventists and Jewish students.

	Church	*Sunday Observance*
Shifts toward	35%	19%
Shifts against	14%	52%

Two and one-half times as many ex-students shifted *toward* the church as *against* it. But two and one-half times as many ex-students shifted *against* Sunday observance as *toward* it. Perhaps the marked shift toward attitudes less favorable to Sunday observance may be related to social changes involving increased tolerance for nonreligious Sunday activities such as secular employment and commercial recreation. However, the intercorrelation between the two tests in 1950 remains positive and significant, .36, *SE* .023.

While significant shifts have been indicated, we find evidence of attitude persistence as shown in the test-retest correlation, .43, *SE* .03. It seems that attitudes measured in college are related to attitudes similarly tested 14 years later.

Types of Institutions

That institutional types differ in student attitudes toward Sunday observance is shown in Table 4. Students at denominational colleges in 1936 were significantly more favorable toward Sunday observance than were those at state universities. Fourteen years later ex-students from all three types of institutions had shifted toward attitudes less favorable to Sunday observance. While ex-students from state universities showed a smaller degree of change toward disapproval of Sunday observance, those from denominational colleges still remained slightly favorable toward Sunday. Compared with denominational ex-students, more individuals from state universities shifted toward Sunday observance and less shifted toward the antagonistic end of the scale.

Regional Differences

As seen by their attitudes toward the church, ex-students tend to retain the regional marks of their collegiate nurture. Those who attended institutions in the South continue more favorable toward Sunday observance than those from institutions in the Midwest and the East.

The Sex Variable

While in college, the 454 women students showed more favor toward Sunday observance than did the 439 men. Fourteen years later, the difference in mean scores had become negligible. However, both sexes had shifted toward more liberal attitudes concerning Sunday observances and these shifts were significant at .01 levels.

TABLE 4

ATTITUDES TOWARD SUNDAY OBSERVANCE BY TYPE OF INSTITUTION

(N=893)

Types of Institution	*N*	1936		1950		Diff.	*p*	No Shift %	Shift Pro %	Shift Anti %
		Mean	*SE*	Mean	*SE*					
4 state univers.	190	5.23	.10	4.84	.13	.39	.01	34	24	42
6 Luth.	359	6.22	.08	5.5	.08	.72	.01	29	19	52
6 other	344	6.00	.08	5.22	.08	.78	.01	27	18	55

Attitudes Toward Sunday and Overt Behavior

We have seen that college attitudes toward the church are related to overt behavior 14 years later as shown by church attendance and tithing. It has also been shown that college tests of attitudes toward Sunday observance are related to retest scores 14 years later ($r = .43$). But are college attitudes toward Sunday related to overt behavior years later? To answer this question, we turn to the correlation between college test scores and participation in church activities. No attempt was made to evaluate or weight such activities. We simply enumerate activities, such as teaching Sunday school, church attendance, service as councilman, or activity in church organizations for men or for women. Between test scores in college and the number of activities in which the ex-student participates 14 years later, we find a small but positive correlation, $r = .34$, *SE* .03. Apparently, attitudes toward Sunday measured in college are related to the number of church activities in which the ex-student engages 14 years later.

V. ATTITUDES TOWARD GOD, A REALITY

Degree of Attitude Change

Since ex-students 14 years after college testing had changed their attitudes significantly in the prochurch direction, it might be reasoned that this would be accompanied by stronger belief in the reality of God. Support for such an hypothesis is given in Table 5. Both 1936 and 1950 means are well above the midpoint of the scale. This means approving items such as: "I trust in God to support the right and condemn the wrong"; "My faith in God develops with experience"; "I am quite convinced of the reality of God." But by 1950 these ex-students had moved in the direction of attitudes still more favorable toward God. The difference between 1936-1950 means, .21, is significant at the .01 level.

Attitude Shifts

The tendency toward more theistic attitudes is evident not only from the increased magnitude of the 1950 mean, but also from the number and percentage of ex-students who had shifted toward the theistic end of the scale contrasted with those moving toward atheism. There were 331 ex-students who shifted to attitudes more favorable to a belief in God than they had while in college. During this 14-year period these same *S*s shifted significantly toward more liberal views in social and economic

TABLE 5

ATTITUDES TOWARD GOD AS A REALITY AMONG STUDENTS AT 16 COLLEGES AND UNIVERSITIES COMPARED WITH THOSE OF THE SAME PERSONS 14 YEARS LATER

($N = 863$)

1936		1950	
Score	*f*	Score	*f*
10	14	10	7
9	151	9	186
8	358	8	431
7	221	7	175
6	77	6	24
5	14	5	16
4	12	4	13
3	11	3	7
2	2	2	3
1	3	1	1
Mean	8.06	Mean	8.27
SE	.045	*SE*	.039

Diff.	.21
p	.01
r	.22

Shifts	*N*	%
No Shift	333	38
Toward God	331	38
Away from God	199	24

TABLE 6

ATTITUDES TOWARD REALITY OF GOD BY TYPES OF INSTITUTIONS ($N=863$)

Types of Institution	N	1936		1950		Diff.	p	No Shift %	Shift Pro %	Shift Anti %
		Mean	*SE*	Mean	*SE*					
4 state univers.	177	7.71	.09	8.02	.08	.31	.01	36	41	23
6 Luth.	356	8.21	.06	8.36	.06	.15	.05	36	37	27
6 other	329	8.09	.06	8.33	.06	.24	.01	43	37	20

areas, as contrasted with the 38 per cent shifting toward belief in reality of God (6). Moving in the opposite or atheistic direction, we find only 24 per cent of the entire group. Approximately 38 per cent retained almost the same position held in college, further evidence of the persistence of college religious attitudes. Persistence of theistic attitudes is further indicated by the correlation, $r = .22$, between 1936 test and 1950 retest scores.

Types of Institutions

When our *S*s are analyzed by type of institutions attended, we find students attending denominational colleges significantly more favorable toward belief in reality of God than those who attended the four state universities (see Table 6). But the 1936 differences between Lutheran and the other denominational institutions were not significant. Ex-students from the "other" denominational schools and from the state universities had increased theistic beliefs by magnitudes significant at the .01 level. After 14 years, however, ex-students from Lutheran colleges had increased their means by a margin significant only at the .05 level. From the six "other denominations" came the largest percentage of "standpatters," 43 per cent had not shifted by so much as one scale point, in contrast with 36 per cent of those who had attended state universities or Lutheran schools. While the greatest mean change, 1936 to 1950, and the greatest percentage of shifts toward theism were among those who had attended state universities, their 1950 mean did not reach that of the other two groups.

Regional Differences

While in college, *S*s attending institutions in the South had mean scores significantly more favorable toward God, a Reality, than had their colleagues from the North and the Midwest. On the retest in 1950, the attitude differences were still in evidence, in the same direction, but to a lesser degree.

The Sex Variable

While the study of 3,649 college students in 1936 revealed a significant sex difference in attitude toward God, Reality, .23 ($p < .01$), with females indicating the stronger belief in God, this was less evident in the 1950 retest when women were only slightly higher than men, 8.29 compared with 8.25.

VI. ATTITUDES TOWARD GOD, AN INFLUENCE ON CONDUCT

We have seen evidence of a shift in attitudes toward the church and a somewhat stronger belief in the reality of God. But do students believe God influences their own personal conduct? Do such beliefs persist after college? Data

TABLE 7

ATTITUDES TOWARD GOD AS AN INFLUENCE ON CONDUCT AT 16 COLLEGES AND UNIVERSITIES COMPARED WITH THOSE OF THE SAME SUBJECTS 14 YEARS LATER

(N=851)

1936		1950	
Score	*f*	Score	*f*
10	85	10	81
9	271	9	333
8	185	8	170
7	110	7	94
6	108	6	100
5	48	5	38
4	14	4	11
3	20	3	15
2	8	2	7
1	2	1	2
Mean	8.21	Mean	8.39
SE	.05	*SE*	.05

Diff.	.18
p	.01
r	.49

Shifts	*N*	%
No shift	272	32
Toward God	319	37
Away from God	260	31

from tests indicating attitudes toward such intimately personal questions are summarized in Table 7.

Degree of Change

That belief in the reality of God would be accompanied by a conviction that God influences conduct might be expected. Such a hypothesis is indeed supported by the correlation between God-reality attitudes and God-conduct, r = .58. A marked similarity is also evident in the degree and pattern of 1936-1950 changes. For the 851 students in 1936, the mean, 8.21, indicates attitudes favorable toward the idea of God influencing conduct. By 1950, the God-conduct concept was viewed even more favorably, a change to a mean of 8.39. The 1936-1950 difference of .18 is significant at the .01 level (see Table 7).

Attitude Shifts

Of the four measures of religious attitudes, we find on the God-conduct scale the highest correlation between scores in college and those of 14 years later, r = .49. Further evidence of persistence over the 14 years is noted in Table 7. There were 272 ex-students, nearly one-third of the 851 tested, who had not shifted by so much as one scale point. Of those who did shift, a majority shifted toward the God-conduct end of the scale, 37 per cent compared with 31 per cent who shifted in the opposite direction.

Types of Institutions

On the God-conduct scale, institutional differences among collegians were evident in 1936. Students at Luthern col-

TABLE 8

ATTITUDES TOWARD GOD AS AN INFLUENCE ON CONDUCT AT DIFFERENT TYPES OF INSTITUTIONS

(N=851)

Types of Institution	*N*	1936		1950		Diff.	*p*	No Shift %	Shift Pro %	Shift Anti %
		Mean	*SE*	Mean	*SE*					
4 state univers.	177	7.46	.12	8.0	.11	.54	.01	26	46	28
6 Luth.	352	8.42	.08	8.45	.07	.03	*	30	36	34
6 other	322	8.38	.09	8.56	.09	.18	*	37	35	28

* Not significant.

leges and at "other" denominational schools showed significantly higher means than did those at state universities (see Table 8). But 14 years later, ex-students from state universities had increased their means by a difference of .54 ($p =$ .01) in the pro-God-conduct direction. While the two groups of denominational colleges changed in the same direction, in neither case was the 1936-1950 difference significant. As was the case with God-reality, we find that on God-conduct a larger percentage of denominational ex-students tend to retain their original scores than do those from state universities. Among denominational students, 30 to 37 per cent of ex-students showed no shift on retest, compared with only 26 per cent of ex-university students.

When we consider percentages of individuals shifting toward the God-concept, state university ex-students lead with 46 per cent compared with only 35 and 36 per cent in the two other groups. It should be noted, however, that while the ex-university students were the only group to shift significantly toward the religious end of this scale, they did not overtake the other two groups.

Regional Differences

On initial test, regional differences follow the pattern already noted for God-reality and church. On retest, ex-students from Southern institutions remained significantly more favorable to God as an influence on conduct than did those from northern colleges.

The Sex Variable

As a group, the 438 women scored higher than men during college and also on the postcollege retest. But the 413 males shifted farther toward the God-conduct belief, hence the 1950 sex difference decreased.

Attitudes Toward God-Conduct and Overt Behavior

While in college these *S*'s indicated belief in God as an influence on conduct, and such belief has persisted through the 14 years following the college tests. But is there any relationship between the belief in college and postcollege conduct?

If belief in God does influence conduct, we might expect evidence from the religious activities in which the ex-student participates. Between college scores on God-conduct attitudes and report of church activities 14 years later, we find a positive correlation ($r = .42$, *SE* .02). Those *S*s who scored high in college tend to be the ones who 14 years later taught Sunday school, participated in church brotherhoods and missionary organizations, and were members of church governing bodies. Those scoring low in college tend to participate in few or none of such church activities.

But what of behavior in community and civic areas? Between college scores on the God-conduct scale and such community activities as participation in community-chest drives, work in welfare organizations, and general civic enterprises, the correlation is positive and significant ($r = .36$, *SE* .03). What a college student believes about the influence of God on conduct is even related, though not strongly, to the percentage of times he votes in state and national elections 14 years later ($r = .23$, *SE* .03).

While not great enough for prediction, all correlations presented are positive and sufficiently large compared with the *SE*'s to sustain the hypothesis that attitudes toward religious values in college are related to report of overt behavior at least 14 years later.

VII. Religious Attitudes and Certain Relationships

Consistency in Religious Attitudes

In three of the four measures of religious attitudes we have considered, certain similarities in pattern have been observed. The changes in means from 1936 test to 1950 retest have been somewhat similar for attitudes toward the church, God-reality, and God-conduct. One might expect attitudes favorable to God-conduct concepts to accompany acceptance of God-reality items. This was supported not only by the pattern of shifts shown previously but also by the correlation, $r = .58$ (see Table 9). We find similar consistency between God-conduct and church, $r = .62$, and between God-reality and church, $r = .51$. These intercorrelations, all above .50 and significant, indicate marked consistency in attitudes. The fact that the correlations are not higher suggests that the tests measure attitudes which are by no means identical. Conceivably, *S*s may accept the reality of God but not go so far as to admit that the existence of God actually influences personal behavior. That the highest intercorrelation ($r = .62$) should be between God-conduct and the church seems not illogical. The purpose of the church is to foster a belief in God. The *S*s who believe God influences conduct might be expected to be favorable to the institution designed to promote godly conduct.

While intercorrelations between God-reality, God-conduct, and the church are all above .50, indicating marked consistency, the correlations between each of these and Sunday observance are significantly lower. Between Sunday observance and church, God-conduct, and God-reality, the *r*'s are .36, .31, .27, respectively. While the latter three apparently have a large factor in common, Sunday observance may involve other factors such as social and economic considerations in addition to the commandment to "remember the Sabbath day. . . ."

Conservatism and Religious Attitudes

Do *S*s who indicate attitudes favorable toward religious values also tend to be conservative on social, economic, and political issues (5, 6)? Is the frequent listing of religion as a "conservative force" an unjustified stereotype (11)? For answers to such questions, we turn to data from the Lentz Opinionnaire, which was included in both test and retest in this study (6). High scores on this test indicate conservatism—low scores show liberal or radical tendency. That *S*s favor-

TABLE 9

Intercorrelations Between Attitudes of Ex-Students*

Attitudes toward	*N*	God-reality	God-conduct	Church	Sunday	*CR*
God-reality	863	—	.58	.51	.27	.24
God-conduct	851	.58	—	.62	.31	.30
Church	887	.51	.62	—	.36	.44
Sunday obs.	893	.27	.31	.36	—	.43
CR	901	.24	.30	.44	.43	—

* All correlations shown are from 6 to 30 times *SE*.

TABLE 10

COMPARISON OF ATTITUDE CHANGES AND NET SHIFTS OVER 14 YEARS

Scale	Direction of change	1936–1950 change in mean scores	1936–1950 Net shifts* %
Church $N=887$	Prochurch	.30	20
God-reality $N=863$	Toward God-reality	.21	14
God-conduct $N=851$	Toward God-conduct	.18	16
Sunday $N=893$	Pro-Sunday observance	−.67	−31
CR $N=901$	Soc.-Pol. conservatism	−1.72†	−20

* "Net Shifts" = % shifting toward a value, less % shifting against.

† The magnitude of the *CR* change is not directly comparable with the change in religious means.

able toward religious values are conservative in other areas would seem apparent from the *CR* column, Table 9. This seems most obvious in correlations between conservatism and the church, $r = .44$, and between conservatism and Sunday observance, $r = .43$. Might relatively high correlation between attitudes toward civic conservatism and attitudes toward the church be explained in terms of the primary function of the church, viz., to perpetuate a heritage of religious values? Perhaps conserving of religious values has spread to other areas as well. The question of including Sunday observance as one of the religious scales has already been raised. Perhaps between Sunday observance and conservatism-radicalism tests there are more factors in common than between Sunday and the other religious scales. Note that between Sunday and political-social conservatism the correlation is $r = .43$, a greater correlation than between Sunday and any of the other religious scales. From the foregoing correlations it would seem that *S*s who score high on religious scales also have a slight tendency to score high on social-economic-political conservatism. But these correlations portray only a static situation at time of testing or retesting. What of the dynamic aspect with which this entire study has been primarily concerned?

A comparison of mean changes in religious attitudes with those in social-political-economic areas is shown in Table 10. For 3 of the 4 religious attitude areas, church, God-reality, and God-conduct, the mean changes were all toward the religious end of the scale and all changes were significant. But for the same *S*s, the mean change in social-political-economic attitudes was toward less conservatism. For the three major areas, net shifts *toward* religious values range from 14 per cent to 20 per cent. On the other hand, for the entire group, the 14 years saw a net shift of 20 per cent away from conservatism, toward liberal attitudes. Again we see Sunday-observance attitude changes and net shifts in the liberal direction, in contrast to the other religious areas.

Attitudes in 1936 and Overt Behavior in 1950

For convenient comparison, correlations between student attitudes in college and various kinds of behavior 14 years

TABLE 11

Attitudes in College and Overt Behavior 14 Years Later

Attitudes in college 1936	Overt behavior 1950	*r*	*SE*
God-conduct	Religious activities	.42	.02
God-conduct	Civic activities	.36	.03
God-conduct	% of time voting	.23	.033
Sunday observance	Religious activities	.34	.03
Church	% Church attendance	.25	.04
Conservative-radical	Civic activities	.012	.026*
Conservative-radical	% of time voting	.05	.05*

* Not significant.

later are summarized in Table 11. It will be noted that college attitudes toward the church, (in 1936) are correlated with church attendance in 1950, $r = .25$, *SE* .04. Attitudes toward Sunday observance expressed in college are correlated with report of participation in religious activities 14 years later, $r = .34$, *SE* .03. Attitudes toward God as an influence on conduct show an even higher correlation with participation in religious activities, $r = .42$, *SE* .02. Thus religious attitudes indicated in college are significantly correlated with religious activities 14 years later. God as an influence on conduct is also related to participation in civic activities, $r = .36$, *SE* .03. This is in sharp contrast to the correlation between conservatism toward political issues expressed in college and participation in civic activities 14 years later, $r = .012$, *SE* .026, not significant. While we found God-conduct attitudes correlated with percentage of time voting, $r = .23$, *SE* .03,conservative-radical test results in college show no such significant correlation with percentage of time voting, $r = .05$, *SE* .05.

Thus it is apparent that college religious attitudes are significantly related to certain types of religious and civic behavior reported 14 years later. On the other hand, college conservatism or liberalism on social-political-economic issues bear no significant relationship to these same activities.

VIII. Summary

1. A general pattern of attitude change toward religion was discernible 14 years after college testing. In three of the four areas studied, retest means were significantly greater than college test means. This proreligious trend was further supported by data on individual shifts: (*a*) Toward the church, 35 per cent came to look with more favor, 14 per cent less. (*b*) Toward God-reality, 38 per cent came to look with more favor, 24 per cent less. (*c*) Toward God-conduct, 37 per cent came to look more favor, 31 per cent less.

2. In sharp contrast, toward Sunday observance, both means and percentages of shifts indicated attitudes significantly less appreciative than while in college.

3. That religious attitudes held in college tend to persist for at least 14 years is supported in all areas of this study. The following evidence was presented: (*a*) Of more than 850 ex-students, 29 per cent to 51 per cent did not show a test-retest attitude change of so much as one scale point. (*b*) Persistence was indicated by the test-retest correlations ranging from .22 to .49 (all correlations significant).

4. The pattern of institutional differences in religious attitudes became less

evident after the 14-year period. Students at state universities in 1936 indicated attitudes less religious than did those at denominational colleges. But ex-students from state universities shifted further toward religious attitudes, so that by 1950 the difference between the two types of institutions was reduced.

5. The regional pattern of 1936 indicated that students in the South were more religious than those of the North and the Midwest. After 14 years, regional differences had declined, with ex-students from the South still indicating the more religious attitudes.

6. On the sex variables, we again see the pattern of decreased differences after 14 years. In 1936 the coeds were more religious than men, but by 1950 sex differences had declined and in no area were they significant at the .01 level.

7. Marked consistency in religious attitudes was indicated by the intercorrelations between God-reality, God-conduct, and the church, all *r*'s above .50.

8. Sunday observance apparently involves less of the religious component common to the other three areas. While intercorrelations between the other three religious areas were all above .50, intercorrelations between Sunday observance and the other three were all below .36. "Net shifts" of attitudes toward Sunday observance were in the direction opposite that of the other three areas.

9. That the attitude toward Sunday observance may be more closely related to conservatism on social-economic-political issues is suggested by a correlation greater than between Sunday observance and each of the other three religious areas.

10. While changes in means, for the most part, were toward proreligious attitudes, in social-economic-political areas the mean changes were toward liberalism.

11. That what an *S* believes in college is related to overt behavior 14 years later was indicated by significant correlations between college attitude scores and such postcollege reports of behavior as church attendance, participation in other religious activities, tithing, civic activities, and percentage of times voting in state and national elections.

REFERENCES

1. Ash, P. Effect of anonymity on attitude questionnaire responses. *J. abnorm. soc. Psychol.,* 1950, 47, 722-723.
2. Ayad, J. M., & Farnsworth, P. R. Shifts in values of opinion items. *J. Psychol.,* 1953, 36, 295-298.
3. Corey, S. M. Signed versus unsigned attitude questionnaires. *J. educ. Psychol.,* 1937, 28, 144-148.
4. Elinson, J., & Haines, Valerie T. Role of anonymity in attitude surveys. *Amer. Psychologist,* 1950, 5, 315. (Abstract)
5. Kirkpatrick, C. Religion and humanitarianism. *Psychol. Monogr.,* 1949, 63, No. 9 (Whole No. 304).
6. Nelson, E. Persistence of attitudes of college students fourteen years later. *Psychol. Monogr.,* 1954, 68, No. 2 (Whole No. 373).
7. Nelson, E. Student attitudes toward college now attended. *Sch. & Soc.,* 1938, 48, 443-444.
8. Nelson, E. Student attitudes toward religion. *Genetic Psychol. Monogr.,* 1940, 22, 323-423.
9. Thurstone, L. L. Scales for measuring attitudes toward religion. Chicago: Univer. of Chicago Press, 1929.
10. Thurstone, L. L., & Chave, E. J. *The measurement of attitude.* Chicago: Univer. of Chicago Press, 1929.
11. Zimmerman, F. K. Religion, a conservative social force. *J. abnorm. soc. Psychol.,* 1934, 28, 473-474.

(Accepted for publication December 5, 1955.)

Vol. 70, No. 18 Whole No. 425, 1956

Psychological Monographs: General and Applied

Political and Religious Dogmatism: An Alternative to the Authoritarian Personality[1]

BUREAU OF EDNL. & PSYL. RESEARCH
LIBRARY

MILTON ROKEACH
Michigan State University

I. INTRODUCTION

THE present report is one in a series on the phenomenon of dogmatism (18, 19, 20, 21, 23). Our objectives here are twofold: (*a*) to present two scales, along with the theory underlying them, for measuring various aspects of dogmatism; and (*b*) to consider several hypotheses regarding the manner in which individual and group differences in dogmatism are related to certain other variables, especially those relevant to present-day thinking and research on the authoritarian personality (1).

Our point of departure in attempting to isolate and measure dogmatism and to test hypotheses relating it to other variables has its roots, on the one hand, in an analysis of certain shortcomings in the formulations by Adorno *et al.* (1) regarding "the authoritarian personality"; and on the other hand, in a theoretical analysis of the properties of dogmatic thinking and believing, and the function it serves for the person. In the succeeding paragraphs we will attempt to spell out more precisely how these two sets of considerations guided both the construction of the measuring instruments and the formulation of specific hypotheses.

We have already suggested elsewhere (19) that the conceptualizations in the *The Authoritarian Personality* fall short of becoming a general theory of authoritarianism and intolerance by virtue of the fact that they had their beginnings in the study of anti-Semitism and ethnocentrism. The California researchers then constructed a personality scale which not only served as an indirect measure of ethnic prejudice but also as a measure of underlying personality predispositions toward an authoritarian outlook on life. This scale was called the

---

[1] I am greatly indebted to the Social Science Research Council for affording me half-time freedom from academic duties over a three-year period in order to carry out the research reported herein. Many persons facilitated this research in various ways. It is possible here to acknowledge my indebtedness to only a few of them. I am grateful to Drs. M. Ray Denny, Al Eglash, Irving Sigel, Henry Clay Smith, and Charles Hanley for their capacity to listen patiently and criticize undogmatically over long periods of time. I am grateful to Dr. Richard Christie of New York University, Dr. Roger W. Russell of University College, London, and Mr. R. R. Hopkins, Personnel and Welfare Manager of Vauxhall Motors in England. They all cooperated well beyond the call of duty in making subjects available to me. I wish to single out especially Miss Joan Ray of University College, London, who gave generously of her time and energy to collect the English data reported herein.

I wish to thank Dr. Cecily de Monchaux and Mr. Arthur Summerfield of University College, London, for their assistance in revising the Opinionation Scale to suit English conditions.

My greatest debt is to my wife, Muriel. More than anyone else she contributed moral support, intellectual stimulation, and original insights into the nature of dogmatic thought and belief.

Bureau Ednl. Psy. Research
DAVID H ING COLLEGE
Dated
Acs No

F scale (F for fascism), indicating that the kind of authoritarianism which the researchers had originally in mind was of the fascist variety. It was only later—perhaps as a result of the publication of the full report of research under the title *The Authoritarian Personality*—that the personality scale referred to came to be widely known as the Authoritarianism scale (24) as well as the F scale.

It was in this way that the general concept, the authoritarian personality, became tied (measured) operationally to the F scale. Both, in turn, became intricately linked, at both theoretical and empirical levels, with anti-Semitism and ethnocentrism and, furthermore, with political-economic conservatism.

Certain consequences follow from the linkage of authoritarianism with the F scale and with anti-Semitism, ethnocentrism, and conservatism. With the more or less rough equating of the general concept of authoritarianism with but one aspect of it (fascist authoritarianism), and by virtue of its theoretical and empirical association with ethnic intolerance, it became somewhat embarrassing to point to manifestations of authoritarianism and intolerance not at all associated or even inversely associated with ethnic prejudice and conservatism.[2] Authoritarianism and intolerance in attitudes and in interpersonal relations can be readily observed among persons adhering to different positions along the total range of the left-to-right continuum—among those left-of-center and center, as well as among those right-of-center—and among persons adhering to different academic orientations in the humanities, the social sciences, and the physical sciences.

The linking together in theory and research of authoritarianism with ethnic prejudice leads to yet other difficulties. It has been frequently observed that there are persons low in prejudice who are high in authoritarianism. It has also been observed that there are persons high in prejudice who are low in authoritarianism. It seems to us that a more general theory of authoritarianism (and the operational measurement thereof) should make it possible to distinguish from among persons relatively lacking in ethnic prejudice those low and those high in authoritarian character structure. Similarly, it should also be possible to distinguish by independent criteria high from low authoritarians among the ethnically prejudiced.

Another obstacle in developing a general theory of authoritarianism is the present-day social climate which leads to a preoccupation with authoritarianism on the left. In tune with the times we live in, a good deal of criticism (4, see especially Shils's chapter) has been leveled against the research by Adorno *et al.* (1), on the ground that it neglected left authoritarianism. This, to our mind, is a legitimate criticism. However, in the present research one of our major purposes is to go beyond this to develop a set of formulations, under the general concept of dogmatism, of the properties of *general* authoritarianism, which is broader than left, right, or center authoritarianism, and which, furthermore, attempts to address itself to manifestations and correlates of authoritarianism in nonpolitical, nonreligious areas as well—the academic world, for instance. In other words, it is our view that a preoccupation with left authoritarianism,[3] while important as such, still leaves one open to the further criticism that what has been neglected is the theoretical analysis and measurement of the properties held in common by all forms of authoritarianism, regardless of specific ideological, philosophical, or scientific content. It is to such broader issues that the theoretical analysis and measurement of dogmatism will attempt to address itself.

---

[2] In this connection the California researchers found it necessary to invent the concept of "the rigid low." At best this is an *ad hoc* concept, necessitating additional *ad hoc* concepts such as "rigid middles," "nonrigid highs," etc.

The problems posed above—regarding the nongeneral character of the research on the authoritarian personality—point strongly to the need to modify, or reformulate, or broaden the already-classical conceptualizations by Adorno *et al.* (1). What direction should such a reformulation take?

1. The analysis of ideological structure. A first requirement, it seemed to us, is the need to make a sharp distinction in theory—and to translate this distinction into empirical research—between the *structure* and *content* of ideological systems. A person may espouse a set of beliefs which are democratic in content; he may take a stand against Jim Crow; he may advocate permissiveness in parent-child relations; he may regard McCarthy as a demagogue. Yet, adherence to democratic beliefs, considered alone, cannot be taken as indicative of a nonauthoritarian outlook. The structure of a belief system must also be taken into account. The problem then becomes to formulate systematically the structural properties of a person's ideology quite apart from its specific content, and in such a way that it can be measured. A successful attack on this problem would open up a number of avenues for further investigation, not the least of which would be the broadening of theory and research on "the authoritarian personality" to extend beyond its present emphasis, which despite its emphasis on underlying personality structure, assumes ideological content as the main criterion of an antidemocratic outlook.

2. The analysis of formal ideological content. Closely related to the structural organization of an ideology is its *formal* content. Two persons differing widely in the content of their religious or political beliefs, for example, may be similar not only in ideological structure but also in formal content. That is, they may both adhere to the belief that there is an Absolute Authority, a True Cause, an Elite, and a Bible. We may say that they differ in specific (substantive) content but are similar in formal content.

3. The analysis of function. In the research on *The Authoritarian Personality* a good deal is said about the psychodynamics underlying adherence to antidemocratic beliefs fascistic in content. In the same way one can inquire into the functions underlying a particular person's ideological structure and formal content, quite apart from specific content.

With these considerations in mind we turned to the phenomenon of ideological dogmatism and to its conceptual representation and measurement. In the previous paper already referred to (19), we have formally defined dogmatism as "(*a*) a relatively closed cognitive organization of beliefs and disbeliefs about reality, (*b*) organized around a central set of beliefs about absolute authority which, in turn, (*c*) provide a framework for patterns of intolerance and qualified tolerance toward others" (p. 195). It is to be noted that according to this definition three sets of variables are subsumed under the construct of dogmatism—closed cognitive systems, general authoritarianism, and general intolerance. The first variable involves an analysis of structural properties; the latter two variables involve an analysis of formal content.

---

[3] A major exception is the stimulating work written by Eric Hoffer (13). In several ways the goals of the present research are similar to those of Hoffer.

In the present monograph we will present two new scales based on this definition—The Dogmatism and Opinionation Scales. In order to show clearly the theoretical assumptions guiding their construction, it will first be necessary to review briefly various ramifications of the above definition as presented in greater detail in the earlier paper (19). This will be supplemented with several further assumptions regarding the function served by dogmatic thinking and believing, which also guided the construction of these scales. Following the presentation of these scales and the theory underlying them (Sections II and III) will be a discussion on "Procedure and Subjects" (Section IV) and a consideration of the reliability and item analysis of the Dogmatism and Opinionation Scales (Section V).

In Section VI we will present the three major hypotheses of this study. Each one of these hypotheses will have a dual purpose: to determine the validity of different aspects of our conceptual definition of dogmatism and by contrast, the extent to which they are compatible or incompatible with the conceptualizations by Adorno *et al.* (1). To anticipate briefly: Hypothesis A is designed to demonstrate that the Dogmatism Scale represents a measure of general authoritarianism, while the F scale represents primarily a measure of right authoritarianism. Hypothesis B is designed to show that dogmatism is related to both left-of-center and right-of-center forms of intolerance, while the F and E scales are constructed in such a way that they tap only right-of-center forms of intolerance. Hypothesis C is to the effect that strong commitment to religious or political ideologies (e.g., the Catholic Church or the Communist Party), will lead to greater dogmatism and opinionation regardless of whether the commitment is left- or right-oriented. In contrast, high F and E scores will be manifest *only* in the case of right-oriented commitments, and not in the case of left-oriented commitment.

II. THE DOGMATISM SCALE

A. THEORETICAL CONSIDERATIONS

1. The Structure of Closed Belief-Disbelief Systems

All cognitive systems, regardless of the extent to which they are open or closed, are conceived as being organized into two interdependent parts: a belief system and a disbelief system. The disbelief system is further conceived as being composed of several disbelief subsystems, rather than just a single one, with each such subsystem arranged along a gradient of similarity to the belief system. We have called this the disbelief gradient. Consider, for example, that we are interested in exploring the ideology of a person whose belief system is Catholicism. We would want to know not only his cognitions and feelings about various aspects of Catholicism, but also his cognitions and feelings about Lutheranism, Calvinism, other Protestant denominations, Judaism, Mohammedanism, etc. Each of the preceding may be said to represent a more or less distinct disbelief subsystem and his knowledge and beliefs about each may well differ from one to the other. It is the task of theory and research to discover the laws governing the organization of these disbelief subsystems with respect to the belief system and with respect to each other. In the earlier paper (19) we advanced a number of properties borrowed largely from Lewinian thought (15), in terms of which belief-disbelief systems are conceived to vary. And in terms of these properties relatively closed belief-disbelief systems are assumed as having a high degree of isolation (17) of the various parts within and between belief and disbelief systems; a relatively greater discrepancy in the degree of differentiation of the belief system as compared with any of the disbelief subsystems; a relatively greater degree of dedifferentiation of the various disbelief subsystems with respect to each other; a relatively greater degree of communication between centrally-located beliefs about absolute authority and peripherally-located beliefs perceived to emanate from such authority; a relatively lower degree of communication (greater isolation) among peripheral beliefs; and a relatively narrower time perspective (8, 15).

2. *The Formal Content of Closed Belief-Disbelief Systems*

We have further assumed that in relatively closed belief-disbelief systems the central core or region has to do with beliefs in absolute authority (both positive and negative) and with related beliefs about "the elite," "the cause," etc. The specific content of such beliefs will, of course, vary with the particular belief-disbelief system adhered to.

A further point is that, coordinated with absolute beliefs in positive and negative authority, are attitudes of intolerance or qualified tolerance toward people in general, depending upon which authorities and belief systems one lines up with, and depending upon which disbelief system one is arrayed against. A major cognitive manifestation of this intolerance is found in the employment of opinionated language (e.g. "Only a simple-minded fool would think that . . . ," "Any intelligent person can plainly see that. . . .")

3. *The Function of Closed Belief-Disbelief Systems*

In the earlier paper (19) we have considered only the structure and formal content of dogmatic thinking and believing. In building our Dogmatism scale, however, it was considered desirable to represent also certain cognitive manifestations of the dynamic aspects of dogmatism.

It is assumed that, whether we are concerned with institutional or individual dogmatism, it represents a total ideological defense against threat and at the same time a cognitive framework for satisfying one's need to know and to comprehend the world one lives in (16). In other words, dogmatic thinking and believing make it possible to ward off threatening aspects of reality and at the same time give one the satisfaction of feeling that one understands it. At the institutional level the threat, whether internal or external, is to the continued existence of the institution and to the belief-disbelief system it stands for. At the individual level the threat arises out of adverse experiences of the person which are shaped by and which, in turn, shape broader social conditions. To varying degrees individuals and groups will become disposed to accept closed systems of thinking and believing in proportion to the extent to which they are made to feel alone or isolated in the world they live in (9), and thus fearfully anxious of what the future holds in store for them. Such a state of affairs leads to pervasive feelings of inadequacy and to self-hate. Attempts are made to overcome such feelings by becoming excessively concerned with a need for power and status. And along with such an overconcern are compensatory attitudes of egotism on the one hand and misanthropy (25) on the other. These, in turn, lead to feelings of guilt and a generally disaffected outlook on life. Such, then, is the constellation of conditions which is conceived to make the individual or group prone to dogmatism. For by overidentification with absolute authority and a cause, an attempt is made to defend self or group against feelings of aloneness and isolation, self-hate and misanthropy. At the same time the closed belief-disbelief system provides a systematic cognitive framework for rationalizing and justifying morally egocentric attitudes toward the self and rejecting attitudes toward others. Thus, the more dogmatic the belief-disbelief system the more does it represent in its totality a tightly-woven network of cognitive defenses against anxiety. Such psychoanalytic defense mechanisms as rationalization, denial, projection, reaction formation, and identification are all seen to have their representation at the cognitive level (as surely they must) in the belief-disbelief system.

B. Construction of the Dogmatism Scale

In constructing the Dogmatism scale it was necessary to assume that the scale would be employed as a research tool primarily in countries where at the very least the word "democracy," whatever it may mean to different individuals or groups, has positive valence. Within this rather broad framework our aim was to devise a measuring instrument generally devoid of content specific to any particular ideological orientation, or having content broad enough so that it was clearly relevant to divergent ideological orientations. This meant that the scale had to be so constructed that individuals adhering dogmatically to such diverse viewpoints as capitalism, communism, Judaism, Catholicism, etc., would all score at one end of the continuum, and opposite in direction to others relatively low in dogmatism.

It is necessary to draw attention further to the fact that the development of the Dogmatism and Opinionation scales and the testing of hypotheses regarding the relation of these scales to other variables proceeded simultaneously. In all a total of 89 items was employed in the initial scale and in four successive revisions covering various facets of the structure, formal content, and function of dogmatism. The several revisions were aimed to increase reliability and to take advantage of modifications and increasing refinements (over a two-year period) in our theoretical formulation of the construct of dogmatism. The initial scale contained 57 items, and three successive editions contained 43, 36, and 66 items, respectively. The fifth and final edition is composed of the best 40 items taken from the preceding 66-item scale, all items significantly differentiating between high and low dogmatism quarters, as determined by item analysis. For the sake of brevity we will present here Form D, composed of 66 items, along with the theoretical considerations guiding its construction.[4] An asterisk preceding the item means that it was retained in the final 40-item scale, Form E.

---

[4] Most of the items in the Dogmatism scale were constructed by the present writer. A few of the items were taken from the work of others, with or without modification. Item 53 was taken from Hoffer (13); Items 46 and 56 from Berger (2); Items 54, 55, 59, 63, 65, and 66 from the MMPI (12).

1. Items Based Upon the Cognitive Structure of Dogmatism

a. Isolation within and between belief and disbelief systems. Isolation refers to degree of segregation or lack of intercommunication between neighboring regions or subregions. It is assumed that the greater the dogmatism the greater the isolation between and within belief and disbelief systems. Cognitive manifestations of isolation and the items they suggested are presented below.

i. Accentuation of differences (or minimization of similarities) between belief and disbelief systems.

*1. The United States and Russia have just about nothing in common.

2. Communism and Catholicism have nothing in common.

3. The principles I have come to believe in are quite different from those believed in by most people.

ii. The perception of irrelevance.

4. In a heated discussion people have a way of bringing up irrelevant issues rather than sticking to the main issue.

iii. The coexistence of contradictions within the belief system.

*5. The highest form of government is a democracy and the highest form of a democracy is a government run by those who are most intelligent.

*6. Even though freedom of speech for all groups is a worthwhile goal, it is unfortunately necessary to restrict the freedom of certain political groups.

7. While the use of force is wrong by and large, it is sometimes the only way possible to advance a noble ideal.

8. Even though I have a lot of faith in the intelligence and wisdom of the common man I must say that the masses behave very stupidly at times.

b. The disbelief gradient. The disbelief gradient is conceived as being composed of a series of disbelief subsystems each arranged along a gradient of similarity to the belief system. Persons adhering to disbelief subsystems most similar to one's own belief system—factional or renegade subsystems—are likely to be in greater conflictful contact with each other and hence are most likely to be perceived as threatening to the validity of the belief system. We assume that this will become increasingly the case the greater the dogmatism.

*9. The worst crime a person could commit is to attack publicly the people who believe in the same thing he does.

*10. In times like these it is often necessary to be more on guard against ideas put out by people or groups in one's own camp than by those in the opposing camp.

*11. A group which tolerates too much differences of opinion among its own members cannot exist for long.

c. Relative degrees of differentiation of belief and disbelief systems. The belief system is assumed to be generally more differentiated than the disbelief system. It is assumed further that with an increase in dogmatism there will be an increasing discrepancy in the degree of differentiation between belief and disbelief systems. Furthermore, the greater the dogmatism, the less differentiation of disbelief subsystems with respect to each other; that is, different disbelief subsystems will be perceived as "the same."

i. Relative amount of knowledge possessed.

*12. It is only natural that a person would have a much better acquaintance with ideas he believes in than with ideas he opposes.

ii. Dedifferentiation within the disbelief system.

13. There are certain "isms" which are really the same even though those who believe in these "isms" try to tell you they are different.

d. Relation between central and peripheral parts. The more closed the belief-disbelief system the more is the central part assumed to correspond to beliefs in and about absolute authority, and the peripheral part to beliefs perceived to emanate from such authority. Thus, a change in a particular peripheral belief is determined by a prior change in the central part. Also the central part is assumed to determine not only what will be represented in the belief-disbelief system but also what will not be represented, that is, narrowed out (16).

i. Tendency to make a party-line change, referring to a change in peripheral beliefs following a change in central beliefs.

*14. In this complicated world of ours the only way we can know what's going on is to rely on leaders or experts who can be trusted.
*15. It is often desirable to reserve judgment about what's going on until one has had a chance to hear the opinions of those one respects.

ii. Narrowing, referring to the selective avoidance of contact with facts, events, etc., incongruent with the belief-disbelief system.

*16. In the long run the best way to live is to pick friends and associates whose tastes and beliefs are the same as one's own.
17. There's no use wasting your money on newspapers which you know in advance is just plain propaganda.
18. Young people should not have too easy access to books which are likely to confuse them.

e. Time perspective. Another dimension along which belief-disbelief systems are conceived to vary is time perspective (8, 15). It is assumed that the more closed the system, the more will its organization be future- or past-oriented, and the more will the present be rejected as important in its own right. Manifestations of such a time perspective are to be found in one's attitude toward past, present, and future, in the extent to which one feels able to appraise accurately or to understand the future, and in one's attitude toward force as a way of revising the present.

i. Attitude toward past, present, and future

*19. The present is all too often full of unhappiness. It is only the future that counts.
20. It is by returning to our glorious and forgotten past that real social progress can be achieved.
21. To achieve the happiness of mankind in the future it is sometimes necessary to put up with injustices in the present.

ii. Knowing the future

*22. If a man is to accomplish his mission in life it is sometimes necessary to gamble "all or nothing at all."
*23. Unfortunately, a good many people with whom I have discussed important social and moral problems don't really understand what's going on.
*24. Most people just don't know what's good for them.
25. There is nothing new under the sun.
26. To one who really takes the trouble to understand the world he lives in, it's an easy matter to predict future events.

iii. Belief in force as a way of revising the present

27. It is sometimes necessary to resort to force to advance an ideal one strongly believes in.

2. *Items Based Upon the Formal Cognitive Content of Dogmatism*

Quite aside from specific ideological content there are assumed to be certain uniformities in the formal content of centrally-located beliefs about the nature of authority and the nature of people in general. With an increase in dogmatism we assume an increase in the absoluteness of beliefs in positive and negative authority, and an increasing rejection or acceptance of people depending upon whether they agree or disagree with one's belief-disbelief system.

a. Authoritarianism

i. Beliefs in positive and negative authority

*28 In the history of mankind there have probably been just a handful of really great thinkers.
*29. There are a number of people I have come to hate because of the things they stand for.

ii. Belief in the cause

*30. A man who does not believe in some great cause has not really lived.
*31. It is only when a person devotes himself to an ideal or cause that life becomes meaningful.
*32. Of all the different philosophies which exist in this world there is probably only one which is correct.
*33. A person who gets enthusiastic about

too many causes is likely to be a pretty "wishy-washy" sort of person.

*34. To compromise with our political opponents is dangerous because it usually leads to the betrayal of our own side.

*35. When it comes to differences of opinion in religion we must be careful not to compromise with those who believe differently from the way we do.

*36. In times like these, a person must be pretty selfish if he considers primarily his own happiness.

37. To compromise with our political opponents is to be guilty of appeasement.

b. Intolerance

*38. There are two kinds of people in this world: those who are for the truth and those who are against the truth.

*39. My blood boils whenever a person stubbornly refuses to admit he's wrong.

*40. A person who thinks primarily of his own happiness is beneath contempt.

*41. Most of the ideas which get printed nowadays aren't worth the paper they are printed on.

42. I sometimes have a tendency to be too critical of the ideas of others.

3. Items Based Upon the Function of Dogmatism

We have already set forth what we consider to be some of the variables having to do with the dynamics underlying dogmatic adherence to any ideology: feelings of aloneness and isolation, anxiety over the future, self-hate, compensatory needs for self-aggrandizement, and a generally paranoid outlook on life. Each of these variables suggested items which appeared to be relevant to the measurement of ideological dogmatism.

a. Fear of aloneness, isolation, and helplessness

*43. Man on his own is a helpless and miserable creature.

*44. Fundamentally, the world we live in is a pretty lonesome place.

*45. Most people just don't give a "damn" for others.

*46. I'd like it if I could find someone who would tell me how to solve my personal problems.

b. Anxiety about the future

i. Fear of the future

*47. It is only natural for a person to be rather fearful of the future.

ii. A feeling of urgency[5]

*48. There is so much to be done and so little time to do it in.

iii. Compulsive repetition of ideas and arguments (self-proselytization).

*49. Once I get wound up in a heated discussion I just can't stop.

*50. In a discussion I often find it necessary to repeat myself several times to make sure I am being understood.

*51. In a heated discussion I generally become so absorbed in what I am going to say that I forget to listen to what the others are saying.

52. In a discussion I sometimes interrupt others too much in my eagerness to put across my own point of view.

c. Self-hate

i. Need for martyrdom

*53. It is better to be a dead hero than to be a live coward.[6]

ii. Conflict within the self.

54. My hardest battles are with myself.

iii. Self-deprecation.

55. At times I think I am no good at all.

56 I am afraid of people who want to find out what I'm really like for fear they'd be disappointed in me.

d. Need for self-aggrandizement

i. Concern with power and status

*57. While I don't like to admit this even to myself, my secret ambition is to become a great man, like Einstein or Beethoven, or Shakespeare.

*58. The main thing in life is for a person to want to do something important.

*59. If given the chance I would do something of great benefit to the world.

60. If I had to choose between happiness and greatness, I'd choose greatness.

ii. Moral self-righteousness

61. It's all too true that people just won't practice what they preach.

e. Paranoid outlook on life

62. Most people are failures and it is the system which is responsible for this.

---

[5] We are indebted to Hoffer (13) for this idea.

[6] In the earlier forms we employed the item, obtained from Hoffer (13), "It is better to be a dead lion than to be a live dog." This item turned out to be highly significant in Form A but not in Form B.

63. I have often felt that strangers were looking at me critically.
64. It is only natural for a person to have a guilty conscience.
65. People say insulting and vulgar things about me.
66. I am sure I am being talked about.

III. The Opinionation Scale

It has already been pointed out in the Introduction that dogmatism is defined not only as a closed cognitive structure of beliefs and disbeliefs organized around a set of centrally-located beliefs about authority, but also as a framework for organizing attitudes of intolerance and qualified tolerance toward people in general according to the beliefs they accept or reject. We have also pointed out that the best single behavioral manifestation of such attitudes of intolerance and qualified tolerance seems to be the employment of opinionated language. *Opinionated rejection* refers to the making of a statement which involves the rejection of a belief and at the same time the rejection of persons accepting this belief. Consider, for example, the following statement: "Only a simple-minded fool would say that God exists." Agreement with this statement implies: (*a*) that the person rejects belief in God and (*b*) that he rejects those who accept belief in God: they are "simple-minded fools."

Half of the items which went into the Opinionation scale are of this type. The principle of construction of such items was simply to link a variety of "opinionated rejection" phrases with a variety of beliefs: about God, socialized medicine, Franco, capitalism, communism, etc.

The remaining half of the items composing the Opinionation scale involve *opinionated acceptance*. This refers to the acceptance of a belief, and along with this the qualified acceptance of persons on condition that and as long as they accept this belief. The following example exemplifies what we have in mind: "Any intelligent person will tell you that God exists." This assertion, too, yields two bits of information: (*a*) the person accepts belief in God and (*b*) he accepts those who agree with him as "intelligent." We say that the acceptance of persons agreeing with the belief is *qualified* because there is the implication that at the moment there is a change of heart the person is no longer considered to be "intelligent."

There is an even more important consideration which guided the construction of the Opinionation scale. On theoretical grounds opinionation should be closely related to dogmatism, and hence a scale for measuring opinionation, as is also the case with dogmatism, should be independent of specific ideological content. To build an opinionation scale without ideological content is obviously impossible. But it is possible to construct a scale having political and religious content so that different ideological positions are more or less equally represented. Accordingly, half of the items were constructed in such a way so that agreement would be indicative of *left opinionation*, while agreement with the remaining half would be indicative of *right opinionation*. "Left" and "right" were defined *not* in terms of beliefs typical of extreme leftist or rightist orientations, but rather in terms of beliefs which could be clearly categorized as "left-of-center" or "right-of-center."

As was the case with the Dogmatism scale, the Opinionation scale also went through several revisions. In all revisions, the scale had the following design: half of the items were *left-opinionation* items, the remaining half right-opinionation items.[7] Each of these, in turn, was composed equally of *opinionated acceptance* items and *opinionated rejection* items. This design make it possible to obtain six different scores, as follows:[8]

1. Total Opinionation (Left Opinionation plus Right Opinionation)
2. Left Opinionation
3. Right Opinionation
4. Opinionated Rejection
5. Opinionated Acceptance
6. Liberalism-Conservatism (Right Opinionation minus Left Opinionation, low [negative] scores indicating liberalism, and high [positive] scores indicating conservatism.

In Table 1 we present the final form of the Opinionation scale, employed in the United States, the 40-item Form C. To be noted is the fourfold classification of items in terms of left and right opinionations, and opinionated acceptance and rejection. To be noted too is the variety of opinionated phrases employed, and the large variety of political and religious issues sampled: Roosevelt, Truman, Hoover, McCarthy, MacArthur, Churchill, Eisenhower, Chiang Kai-Shek, the Rosenbergs, Alger Hiss, Russia, the United States, Spain, China, the United Nations, communism, socialism, capitalism, God, religion, socialized medicine, the Democrats, civil rights, loyalty oaths, congressional committees, the American Legion, labor, race, housing, and military training.

Form C differs from Form C_E (used in England) and from the earlier Forms A and B in certain important respects. This can be discussed more profitably later, in connection with the presentation of the reliability data.

IV. PROCEDURE AND SUBJECTS

The items of the Dogmatism and Opinionation scales were interspersed among items from the several other scales which were also administered. These included the F, E, P.E.C. scales (1) and the Anxiety scale (26). Moreover, to certain groups, the Rigidity (11) and Intellectual Conviction (20) scales were also given.[9] The instructions were the same as those employed in the original research on the authoritarian personality. Also, as in the earlier research, *S*s were asked to indicate disagreement or agreement on a scale from -3 to $+3$, with the 0 point excluded in order to force responses toward disagreement or agreement. This scale was subsequently converted, for scoring purposes, to a 1 to 7 scale by adding a constant of 4 to each item score. The questionnaires were administered to various groups from about 10 to 50 under anonymous conditions, either in regular class periods or at other scheduled times. The questionnaires contained, for different groups, from 175 to 200 items and it took anywhere from half an hour to an

---

[7] To obtain an independent check on whether the items constructed were indeed left or right in direction we asked Drs. Alfred Dietze and Donald M. Johnson to judge each item as to whether it is "left-of-center" or "right-of-center." The agreement of these judgments with the writer's was 100 per cent.

[8] In the interest of brevity we will omit the correlational data on opinionated rejection and opinionated acceptance (scores 4 and 5). Each of these two subscales correlates in the .80's with total opinionation; each correlates about equally highly with other variables. Thus, these two subscales appear to be equivalent halves of the total Opinionation Scale.

[9] Data bearing on anxiety, rigidity, and intellectual conviction are treated elsewhere (20, 21, 23).

TABLE 1

The Opinionation Scale

Left Opinionation	Right Opinionation
Opinionated Rejection	
1. It's just plain stupid to say that it was Franklin Roosevelt who got us into the war.	21. It's simply incredible that anyone should believe that socialized medicine will actually help solve our health problems.
2. A person must be pretty stupid if he still believes in differences between the races.	22. A person must be pretty ignorant if he thinks that Eisenhower is going to let the "big boys" run this country.
3. There are two kinds of people who fought Truman's Fair Deal program: the selfish and the stupid.	23. It's the fellow travellers or Reds who keep yelling all the time about Civil Rights.
4. A person must be pretty short-sighted if he believes that college professors should be forced to take special loyalty oaths.	24. It's the radicals and labor racketeers who yell the loudest about labor's right to strike.
5. It's the people who believe everything they read in the papers who are convinced that Russia is pursuing a ruthless policy of imperialist aggression.	25. It is foolish to think that the Democratic Party is really the party of the common man.
6. It's mainly those who believe the propaganda put out by the real-estate interests who are against a federal slum clearance program.	26. You just can't help but feel sorry for the person who believes that the world could exist without a Creator.
7. A person must be pretty gullible if he really believes that the Communists have actually infiltrated into government or education.	27. It's usually the trouble-makers who talk about government ownership of public utilities.
8. It's mostly those who are itching for a fight who want a universal military training law.	28. Only a misguided idealist would believe that the United States is an imperialist warmonger.
9. It is very foolish to advocate government support of religion.	29. It's mostly the noisy liberals who try to tell us that we will be better off under socialism.
10. Only a simple-minded fool would think that Senator McCarthy is a defender of American democracy.	30. It's the agitators and left-wingers who are trying to get Red China into the United Nations.
Opinionated Acceptance	
11. It's perfectly clear that the decision to execute the Rosenbergs has done us more harm than good.	31. Any intelligent person can plainly see that the real reason America is rearming is to stop aggression.
12. Any person with even a brain in his head knows that it would be dangerous to let our country be run by men like General MacArthur.	32. Plain common sense tells you that prejudice can be removed by education, not legislation.
13. The truth of the matter is this! It is big business which wants to continue the cold war.	33. Anyone who is really for democracy knows very well that the only way for America to head off revolution and civil war in backward countries is to send military aid.
14. Make no mistake about it! The best way to achieve security is for the government to guarantee jobs for all.	34. History will clearly show that Churchill's victory over the Labour Party in 1951 was a step forward for the British people.
15. It's perfectly clear to all decent Americans that Congressional Committees which investigate communism do more harm than good.	35. The American re-armament program is clear and positive proof that we are willing to sacrifice to preserve our freedom.
16. Thoughtful persons know that the American Legion is not really interested in democracy.	36. This much is certain! The only way to defeat tyranny in China is to support Chiang Kai-Shek.
17. It's perfectly clear to all thinking persons that the way to solve our financial problems is by a soak-the-rich tax program.	37. It's already crystal-clear that the United Nations is a failure.
18. It's all too true that the rich are getting richer and the poor are getting poorer.	38. A study of American history clearly shows that it is the American businessman who has contributed most to our society.
19. History clearly shows that it is the private enterprise system which is at the root of depressions and wars.	39. Even a person of average intelligence knows that to defend ourselves against aggression we should welcome all help—including Franco Spain.
20. Anyone who's old enough to remember the Hoover days will tell you that it's a lucky thing Hoover was never re-elected.	40. Anyone who knows what's going on will tell you that Alger Hiss was a traitor who betrayed his country.

hour and a half for completion. Other information, including age, sex, race, religion, political affiliation, and income, was also obtained.

The *S*s were obtained from three areas differing in political climate: the Midwest, New York, and England. In the Midwest the *S*s were college students at Michigan State University taking courses in beginning psychology (M.S.U. I, II, and III). They were tested during the regular class period. In New York, *S*s were also beginning psychology students obtained from two colleges in Metropolitan New York and were also tested during regular class periods (N.Y. Colleges). In England, the writer obtained college samples from University College, London, and Birkbeck College, and a worker sample from a British automobile factory (Vauxhall Motors). The college sample was composed of first-year psychology students recruited from psychology classes and also of individually-recruited volunteers from the college at large (English Colleges I and II). They were tested in small groups from 10 to 15. The worker group (English Workers) was composed of 15 out of 20 workers elected to the Management Advisory Committee (representing 14,000 workers) and 45 more who were individually solicited by the members on this committee. The number of *S*s obtained from all sources, both in the United States and in England, is shown in Table 2.

In view of the fact that the California E scale—we employed the 10-item E scale (1, p. 128)—contains items about Jews and Negroes, questionnaires obtained from Jews and Negroes were eliminated from consideration in the M.S.U. groups. In the New York study we employed a revised 10-item E scale, taken from the longer scale employed in the California research, in which none of the items referred to Jews. By virtue of the fact that Negro items were included, we eliminated Negro *S*s in analyzing these results. In England yet other revisions were necessary. Three E items (1, Items 5, 10, 15 on p. 128) were not relevant and other items were substituted. Also, the word *Britain* was substituted for *America* in the last E item. Jewish *S*s were retained, but all data involving the E scale were analyzed with the Jews omitted from consideration.

V. Reliability and Item Analyses of the Dogmatism and Opinionation Scales

A. The Dogmatism Scale

We have already pointed out that the Dogmatism scale went through a number of revisions. These revisions were made in order to increase reliability and also to reflect the many modifications, elaborations, and refinements which took place in our thinking over a period of two years regarding the nature and meaning of dogmatism. Thus in the various revisions we found it necessary to eliminate certain items, reword some, and add new ones.

Along with the construction of the Dogmatism scale we kept an eye out for its validity. Our approach to the problem of validity was first to analyze certain implications of our conceptualization of dogmatism, translate these implications into hypotheses regarding the way in which dogmatism should be related to other variables, and then to test these hypotheses empirically. The hypotheses and results found will be considered in Sections VI to X.

We will be concerned here with the

Bureau Ednl. Psy. Research
DAVID HA[illegible]AINING COLLEGE
Dated . ...
Accs. No ....

TABLE 2

Corrected Reliability of Successive Forms of the Dogmatism Scale

Form	Number of Items	Group	*N*	Reliability	M	*SD*
A	57	M.S.U. I	202	.70	182.50	26.20
B	43	N.Y. Colleges	207	.75	141.35	27.21
C	36	(N.Y. Colleges	207	.84)*	120.00	19.46
		M.S.U. II	153	.73	126.92	20.14
		M.S.U. III	186	.71	128.34	19.17
D	66	English Colleges I	137	.91	219.10	28.30
E	40	English Colleges II	80	.81	152.80	26.20
		English Workers	60	.78	175.80	26.00

* Based upon a rescoring of the 36 best items out of the 43 items in Form B.

problem of reliability. In Table 2 we show the odd-even reliabilities, corrected by the Spearman-Brown formula, not only for the 66-item Form D but for all forms. Also shown are the number of items in each form, the groups to which each form was administered, number of cases, means and sigmas.

The initial Form A was composed of 57 items and its corrected reliability was .70. Following an item analysis we eliminated 14 items and used Form B, composed of 43 items. Its corrected reliability was .75. A second item analysis led to Form C, which was essentially the same as Form B except that the 7 items with the poorest discriminatory power were further eliminated. Form C, however, yielded reliabilities somewhat lower than we had expected ($r = .73$ and .71), and in the next revision (Form D) we added 30 new items, making a total of 66 items. The corrected reliability of the 66-item Form D was .91. To shorten the scale, we eliminated 26 items on the basis of a third item analysis. This 40-item scale, Form E, was found to have a corrected reliability of .81 for the English College II sample and .78 for the English Worker sample.

The means and sigmas shown in Table 2 are presented primarily for normative purposes. In general, the means for the various groups are not comparable, primarily because the number of items differ for the various forms. One exception is M.S.U. II vs. III, both tested with Form C. It is seen that the means and sigmas for M.S.U. II and III are almost identical. A second exception is English Colleges II vs. English Workers, both tested with Form E. The mean dogmatism score for the English Workers is considerably larger than that for the English Colleges II group, the difference being significant well beyond the one per cent level. It is not possible to say whether this means that the English workers are more dogmatic than English college students or whether the difference is a function of the fact that workers, having considerably less education, agree more often with statements presented to them. The writer is inclined more toward the latter interpretation.

We show in Table 3, for each of the 66 items of Form D, its mean and standard deviation, as well as means and standard deviations, differences in means, and *t* ratios for the high and low dogmatism quarters (upper and lower 27 per cent). As indicated earlier in Section II an asterisk preceding the item means that it went to make up the final 40-item scale, Form E. Also shown, are the *t* ratios for the high and low dogmatism groups for 36 of the 66 items, which had been employed previously in Forms A and B.

To be noted first are the items found to be most discriminating. Considering those items which were employed in the several forms, there were four which yielded an average *t* of 5.00 or more. These are Item 31 (Life meaningful only when there is devotion to cause), Item 43 (Man on his own is helpless and miserable), Item 46 (Want to find someone to solve personal problems), and Item 63 (Strangers look at me critically).

Items employed for the first time in Form D, which yielded a *t* of 6.00 or more, are: Item 38 (2 kinds of people: for truth, against truth), Item 47 (Natural to fear future), and Item 16 (Pick friends whose tastes and beliefs are same as one's own).

Furthermore, we single out the following as also being among the most discriminating (*t* of 5.00 or more for items in Form D only; *t* of 4.00 or more for items used in the several forms):

9. Worst crime is to attack those of similar beliefs.
10. Be more on guard against subversion from within than from without.
11. Groups tolerating diverse opinions cannot exist for long.
14. To know what's going on, rely on trusted leaders.
23. Most people don't understand what's going on.
28. In all history there is just a handful of really great thinkers.
39. My blood boils when others won't admit they're wrong.
59. If given chance I'd benefit world.
65. People say insulting things about me.

A word about the final Dogmatism scale, Form E: it will be noted that 21 out of the 40 items of Form E came from among the 30 items employed for the first time in Form D, while 19 came from earlier forms. There were several criteria which served as bases for selection: (*a*) Items employed for the first time in Form D were eliminated if they discriminated relatively poorly. (*b*) Items employed in the several forms were eliminated if they did not discriminate consistently well. (*c*) The final scale should have approximately equal numbers of items having to do with the structure, the formal content, and the function of dogmatism. (*d*) It should have maximum theoretical coverage. (*e*) Finally, it was our judgment that, quite aside from the issue of external validity, each of the scale items should have "face validity," that is, its content should be directly relevant to our conceptual definition of dogmatism. This necessitated excluding several statistically excellent items (see especially Items 63 and 65) from the final scale which, in our judgment, did not have such "face validity."

B. The Opinionation Scale

In Table 4 we show the corrected reliabilities of the successive forms of the total Opinionation scale, the left- and right-opinionation subscales, and the derived liberalism-conservatism score, obtained by subtracting the left-opinionation score from the right-opinionation score.

Form A, the initial form, was composed of 40 items, broken down into 20 pairs of statements. Each pair was composed of a left-opinionated statement and an oppositely-worded right-opinionated statement. Five pairs were presented together on a single page to *S*, and in all there were four such pages interspersed among other pages containing items measuring other variables. In responding

TABLE 3

Item Analysis of the Dogmatism Scale, Form D

Items	Total Group (N=137)		Highs (N=37)		Lows (N=37)		Diff.	t^a	Form A t	Form B t
	M	σ	M	σ	M	σ				
A. Based upon the structure of dogmatism.										
*1. U.S. and Russia have nothing in common.	2.88	1.77	3.43	1.94	2.16	1.44	1.27	3.19	2.36	3.88
2. Communism & Catholicism have nothing in common.	3.54	2.06	4.32	2.17	2.68	1.72	1.64	3.61	3.84	1.50
3. My principles are different from most others.	3.51	1.97	3.70	1.90	3.14	1.88	.56	N.S.	—[b]	—
4. People bring up irrelevant issues.	5.99	1.39	6.46	.64	5.81	1.43	.65	N.S.	—	—
*5. Belief in democracy run by most intelligent.	4.72	2.03	5.60	1.93	3.92	2.15	1.68	3.54	—	—
*6. Belief in free speech, but not for all.	4.16	2.08	5.11	2.02	3.27	1.85	1.84	4.06	—	—
7. Force is wrong, but sometimes necessary.	3.89	2.06	4.54	2.03	3.51	2.11	1.03	2.14	—	—
8. Masses intelligent, but also stupid.	5.64	1.48	6.11	1.03	5.00	1.86	1.11	3.18	—	—
*9. Worst crime is to attack those of similar beliefs.	2.74	1.69	3.30	1.87	1.76	.84	1.54	4.57	4.37	5.46
*10. Guard against subversion from within.	3.96	1.78	4.78	1.49	2.81	1.62	1.97	5.44	—	—
*11. Groups tolerating diverse opinions can't exist.	3.35	2.01	4.32	1.93	2.19	1.57	2.13	5.22	—	—
*12. Better knowledge of beliefs than disbeliefs.	4.99	1.82	5.65	1.47	4.14	2.10	1.51	3.58	—	—
13. Certain "isms" really the same, not different.	4.59	2.01	5.19	1.78	4.08	2.13	1.11	2.42	—	—
*14. To know what's going on, rely on leaders.	3.80	2.14	5.46	1.61	3.08	1.96	2.38	5.68	—	—
*15. Reserve judgment until you hear leader's opinion.	5.20	1.93	5.70	1.43	4.32	1.92	1.38	3.51	—	—
*16. Pick friends who believe as you do.	3.98	1.98	5.08	1.70	2.68	1.72	2.40	6.04	—	—
17. Don't waste money on reading opposing views.	3.42	1.94	3.76	2.00	2.97	1.86	.79	N.S.	—	—
18. Keep young people away from confusing books.	3.02	1.54	3.65	2.29	2.76	1.90	.89	N.S.	—	—
*19. Present unhappy. The future is what counts.	3.15	1.90	3.81	1.86	2.35	1.51	1.46	3.71	4.58	5.82
20. To progress, return to glorious past.	1.58	.98	1.73	.94	1.38	.67	.35	N.S.	5.23	4.35
21. For happiness in future, present injustice necessary.	4.57	1.75	5.35	1.48	4.22	1.89	1.13	2.87	1.67	4.08
*22. To accomplish mission, gamble all or nothing.	5.04	1.82	5.84	1.35	4.35	1.93	1.49	3.41	2.10	5.46
*23. Most people don't understand what's going on.	4.70	1.77	5.43	1.46	3.51	1.84	1.92	4.97	4.27	5.29
*24. Most people don't know what's good for them.	4.10	2.12	5.16	1.82	2.81	2.22	2.35	4.97	—	—
25. Nothing new under the sun.	3.12	2.12	3.81	2.30	2.38	1.81	1.43	2.98	5.22	2.81
26. If you understand, it's easy to predict future.	2.72	1.78	3.19	2.00	2.27	1.59	.92	2.19	3.13	3.21
27. Force necessary to advance ideal.	3.71	2.00	4.43	1.99	3.11	2.05	1.32	2.81	4.81	4.86
B. Based upon formal content of dogmatism.										
*28. Just a handful of great thinkers.	4.85	2.05	5.87	1.53	3.60	2.03	2.27	5.42	—	—
*29. I hate some people because of what they stand for.	3.07	2.14	3.35	2.21	1.95	1.80	1.40	3.00	—	—
*30. A man without a cause hasn't lived.	4.52	2.09	5.51	1.74	3.27	1.99	2.24	5.12	5.57	3.88
*31. Life meaningful when there is devotion to cause.	5.34	1.66	6.14	.89	4.19	2.00	1.95	5.42	2.70	8.60
*32. There is only one correct philosophy.	2.66	2.23	4.11	2.54	1.97	1.70	2.14	4.24	—	—
*33. Person believing in too many causes is "wishy-washy."	3.20	1.95	4.32	1.76	2.54	1.44	1.78	4.78	—	—
*34. To compromise is to betray own side.	2.78	1.59	3.32	1.71	1.95	1.06	1.37	4.11	2.59	5.59
*35. In religion, we should not compromise.	2.58	1.87	3.46	2.27	1.92	1.26	1.54	3.60	4.36	3.51
*36. To consider only one's own happiness is selfish.	4.21	2.01	5.41	1.44	3.62	2.93	1.79	3.31	3.95	4.50
37. To compromise is to appease.	2.23	1.48	2.95	2.02	1.70	.83	1.25	2.28	5.62	4.10
*38. Two kinds of people: those for, those against truth.	2.58	1.53	3.97	1.99	1.57	.86	2.40	6.74	—	—
*39. My blood boils when others won't admit they're wrong.	4.66	1.76	5.30	1.64	3.87	1.24	1.43	4.22	4.51	6.06
*40. One who thinks of own happiness beneath contempt.	4.17	1.93	4.97	1.71	3.57	1.88	1.40	3.35	3.44	5.41
*41. Most printed ideas aren't worth paper printed on.	3.52	1.94	4.73	1.91	2.73	1.75	2.00	4.68	—	—
42. I'm too critical of others' ideas.	4.81	1.73	5.30	1.53	4.19	1.89	1.11	2.77	—	—
C. Based upon the function of dogmatism										
*43. Man on his own is helpless and miserable.	3.43	2.18	4.92	2.08	2.41	1.81	2.51	5.55	4.14	5.82
*44. World a lonesome place.	3.23	1.96	3.89	2.02	2.41	1.64	1.48	3.47	—	—
*45. Most people don't give a "damn" for others.	3.75	1.87	4.57	1.81	2.81	1.68	1.76	4.33	—	—

TABLE 3 (*Continued*)

Items	Total Group (*N* = 137)		Highs (*N* = 37)		Lows (*N* = 37)		Diff.	t^a	Form A t	Form B t
	M	σ	M	σ	M	σ				
*46. I want to find someone to solve my problems.	3.61	2.22	4.43	2.10	2.76	1.99	1.67	3.52	6.50	6.19
*47. Natural to fear future.	3.79	1.81	4.81	1.61	2.57	1.84	2.24	6.10	—	—
*48. So much to do, so little time to do it in.	5.86	1.69	6.49	1.01	5.19	2.06	1.30	3.44	—	—
*49. Once I get wound up, I can't stop.	3.34	1.94	3.73	2.01	2.43	1.37	1.30	3.24	3.77	5.17
*50. I repeat myself to make sure I'm understood.	3.50	1.91	4.30	1.83	2.78	1.72	1.52	3.65	—	—
*51. I don't listen.	3.44	1.80	3.92	1.87	2.78	1.71	1.14	2.73	4.63	5.55
52. I interrupt others to put across my own views.	4.43	1.85	4.78	1.68	3.65	2.00	1.14	2.64	—	—
*53. Better be dead hero than live coward.	3.26	1.94	4.05	1.94	2.60	1.62	1.45	3.51	—	3.03
54. Hardest battles are with myself.	4.70	2.01	5.38	1.67	3.84	1.76	1.54	3.86	4.20	6.19
55. I'm no good.	4.60	2.02	5.38	1.67	3.84	2.08	1.54	3.51	5.03	6.14
56. I'm afraid people will find out what I'm really like.	2.83	2.03	3.49	2.24	1.97	1.54	1.52	3.38	5.50	4.56
*57. Secret ambition is to become a great man.	5.02	2.05	5.97	1.52	4.20	2.37	1.77	3.84	4.36	5.49
*58. Main thing in life is to do something important.	4.22	2.04	5.38	1.96	3.32	2.18	2.06	4.26	—	—
*59. If given chance I'd benefit world.	5.51	1.63	6.43	.68	4.81	1.86	1.62	4.96	4.00	5.58
60. Greatness more important than happiness.	2.05	1.74	2.89	2.34	1.97	1.65	.92	N.S.	4.54	3.72
61. People won't practice what they preach.	5.55	1.36	5.97	1.15	5.16	1.65	.81	N.S.	1.89	3.63
62. Most people failures and the system is responsible.	3.31	1.96	3.62	2.16	2.70	1.86	.92	N.S.	4.19	4.67
63. Strangers look at me critically.	4.10	1.94	5.22	1.69	3.35	1.91	1.87	4.44	5.09	6.00
64. Only natural to have guilty conscience.	3.35	2.04	4.65	1.92	2.46	1.68	2.19	5.38	3.58	4.78
65. People say insulting things about me.	2.68	1.71	3.78	1.86	2.03	1.48	1.75	4.48	4.48	5.61
66. I'm talked about.	3.25	1.99	4.22	1.91	2.76	1.89	1.46	3.29	6.64	4.69

[a] Employing a one-tailed t test, t's of 1.64 and 2.33 are significant beyond the 5 and 1 per cent levels, respectively, for large samples.

[b] A dash (—) means that the item in question was devised for a subsequent form.

* Item retained in final 40-item scale.

to Form A, *Ss* employed a +3 to −3 agree-disagree scale with the 0 point excluded. As can be seen from Table 4 the corrected reliability of the total scale is .67 and .75. Of special interest is that the two subscales (L.O. and R.O.), despite the fact that they are only half as long as the total opinionation scale, have reliabilities as high as the total scale. The reliability of the liberalism-conservative score (R.O. minus L.O.) is also as high or higher.

Form B was composed of the best 16 (out of 20) pairs taken from Form A, as

TABLE 4

CORRECTED RELIABILITIES OF THE TOTAL OPINIONATION SCALE (T.O.), LEFT OPINIONATION (L.O.), RIGHT OPINIONATION (R.O.), AND LIBERALISM-CONSERVATISM (R.O.-L.O.)

Form	Number of Items	Group	*N*	T.O.	L.O.	R.O.	R.O.-L.O.	Correlation between L.O. & R.O
A	40	M.S.U. I N.Y. Colleges	202 207	.67 .75	.64 .74	.67 .77	.66 .83	−.22 −.51
B	32	M.S.U. II	153	.68	.39	.68	.50	.09
C	40	M.S.U. III	186	.76	.68	.77	.70	.00
C_E	40	English Colleges I English Colleges II English Workers	137 80 60	.75 .75 .75	.89 .90 .91	.88 .86 .91	.93 .93 .94	−.65 −.61 −.62

determined by item analysis. All 16 pairs were presented consecutively (without interspersion with other items) and *S*s marked each item in terms of the following qualitative scale:

1. I disagree with the statement
2. I agree with the statement in part
3. I agree with the statement

The purpose of this 3-point qualitative scale was to give *S* the opportunity to agree with the content of the statement and at the same time to indicate, if he so desired, his disapproval or rejection of the opinionated manner in which the statement is put. It was reasoned that a person agreeing both with the content of the statement *and* the opinionated manner in which it is put would respond with a "3," while another person agreeing with the content but not with the opinionated manner in which it is put would prefer to respond with a "2." Thus, only "3" responses can be assumed to indicate "opinionation," and the total opinionation score is simply the number of statements responded to with a "3." (Responses of the "2"-type indicate "nonopinionation," and "1"-type responses are indeterminate with respect to opinionation since *S* presumably disagrees with the content of the statement. It should be noted, furthermore, that the three possible responses represent a nominal scale because they are qualitatively different from each other and not linear, in contrast to the 6-point Likert-type scale employed in Form A, which is linear.)

It can be seen from Table 4 that this change in rating procedure (Form B) did not raise the reliability of the total scale, as compared with Form A, and in fact lowered substantially the reliability of the left opinionation ($r = .39$) and liberalism-conservatism scores ($r = .50$).

In a further attempt to increase reliability we developed Forms C (presented in Table 1) and C_E. Form C_E was employed in England. It is identical in structural design to Form C with the exception that its wording and content were altered wherever necessary to suit British conditions.[10] Both forms are composed of 40 items, and once again responses were on a -3 to $+3$ scale with the 0 point excluded. In contrast to the preceding forms A and B, the 20 left- and the 20 right-opinionation items are not paired or matched in content, that is, they are highly diversified in content. A major reason for abandoning the matched left and right pairs is that we suspected that *S*s perceived the pairs as opposites and thus felt that they had to agree with one of the items, and disagree with its opposite. Thus it was possible to intersperse each item among the many other items of the questionnaire designed to measure other variables.

It is seen that this revision did not affect the reliability of the total Opinionation scale. In fact, it is noteworthy that despite the various revisions in wording and item content, in pairing or nonpairing of items, in extent of interspersion with other items, and in type of agree-disagree scale employed, that the reliability of the total scale remains about the same ($r = .67$ to $.75$). Moreover, the reliabilities of the Left- and Right-Opinionation subscales are about the same as those for Form A, in which the left and right statements were paired and presented together. Similarly, the reliabilities of the Left- and Right-Opinionation subscales and of the liberalism-conservatism measure are as high as the reliability of the total scale. And with Form C_E we note that the reliabilities of the two subscales and the liberalism-conservatism measure are consistently *higher* ($r = .86$-$.94$) than those found for the total scale ($r = .75$). At first glance, these findings seem suspect in that they appear to violate the widely-accepted notion that subscale reliabilities cannot be as high or higher than reliabilities of total scales. These unusual findings are at least partly explained by the fact that in general left and right opinionation correlate either negligibly or negatively, as shown in the last column of Table 4. Thus, it is seen that the total Opinionation scale, by virtue of the logic of its design, is inherently more heterogeneous than either the Left- or Right-Opinionation subscales. Moreover, these findings suggest that further efforts to increase the reliability of the total scale, given a test of about 40 items, would probably prove unrewarding.

---

[10] A copy of Form C_E is available upon request.

Let us consider further the correlation between left and right opinionation for the various groups. We note that for the three Michigan State University groups the correlations are −.22, .09, and .00, respectively. For the New York Colleges group, the correlation is −.51, and for the three English groups the correlations are −.65, −.61, and −.62, respectively. These findings strongly suggest that there are sharp discrepancies among the three sets of groups in the organization of attitudes along the left-right dimension. For the Michigan State University groups, *S*s are about as likely as not to adhere with varying degrees of opinionation to both left and right beliefs at the same time. That is, it is not possible to predict scores on left opinionation from scores on right opinionation. In contrast, the negative correlation of −.51 between left and right opinionation for the New York college group indicates a more consistent organization of political and religious beliefs in terms of the left-to-right dimension. And this trend is even more apparent in the two English college groups and in the English worker group.

The findings with respect to left and right opinionation parallel the reliabilities of the liberalism-conservatism score. The highest reliabilities are found for the three English groups (.93, .93, and .94), the next highest is found for the New York group (.83), and the lowest are found for the three Michigan State University groups (.66, .50, and .70).

It may be suggested that the findings discussed above are relevant to the study of regional or national character and of modal personality (14). Further research regarding attitude organization in various regions of the United States and in various foreign countries, and a careful analysis of the implications of such attitude organization for intergroup relations, might prove rewarding.

Consider next the means and standard deviations on total opinionation, left opinionation, and right opinionation, as shown in Table 5. It is seen that for all American groups the mean right-opinionation scores are considerably higher than the mean left-opinionation scores (and significantly so). The same is true for the English worker group (but not significantly). The reverse is the case for the two English college groups, whose mean left-opinionation scores are significantly higher than the mean right-opinionation scores.

To be noted also is that the mean score on total opinionation remains approximately the same for all groups tested with Forms A, C, and C_E (which are roughly but not completely comparable). The only exception is the English worker group, which scores significantly higher than each of the remaining groups (excluding from consideration the M.S.U.

TABLE 5

MEANS AND STANDARD DEVIATIONS ON LEFT, RIGHT, AND TOTAL OPINIONATION

Form	Number of Items	Group	N	Left Opinionation		Right Opinionation		Total Opinionation	
				M	σ	M	σ	M	σ
A	40	M.S.U. I	202	59.20	11.95	87.25	13.75	146.45	16.03
		N.Y. Colleges	207	66.67	17.28	79.61	18.39	146.28	17.69
B[a]	32	M.S.U. II	153	1.16	1.32	5.75	3.13	6.91	3.56
C	40	M.S.U. III	186	61.17	11.94	80.84	14.60	142.01	18.94
C_E	40	English Colleges I	137	77.77	21.80	67.23	21.30	145.00	17.80
		English Colleges II	80	76.60	23.00	62.55	17.90	139.15	17.90
		English Workers	60	75.40	25.50	80.85	25.70	156.25	19.40

[a] See text for a discussion of the low means for Form B.

II group). This finding is in agreement with the findings reported earlier regarding the higher mean dogmatism score of the English Worker group (Table 2). And as already stated, it is not possible to decide whether the higher total opinionation mean of this group indicates that they are indeed more opinionated or, more simply, that they have a greater tendency, due to lesser education, to endorse the statements presented.

We turn next to consider the results of the item analysis of the Opinionation scale, Form C. These are shown in Table 6. Also shown are the *t* ratios for 20 of the 40 items which had been employed in Form A.[11]

There are a number of points which should be noted:

1. It is a basic principle of test construction that double-barreled statements should be avoided. This principle is necessarily violated in measuring opinionation because by definition this variable involves the frequency with which individuals accept or reject such double-barreled statements. This is bound to affect both the reliability of the total scale and the discriminatory power of each item. It is therefore all the more surprising that the Opinionation scale and subscales achieve the relatively high reliabilities that they do, and that the great majority of the items (37 out of 40 on Form C) significantly differentiate upper from lower opinionation quarters beyond the 5 per cent confidence level.

2. The right-opinionation items generally discriminate more sharply than the left-opinionation items.

3. On the whole the difference in means on the various items between high- and low-opinionation group is generally lower than that between high- and low-dogmatism groups for the items on the Dogmatism scale. This is probably due to the fact, as already noted, that the Opinionation scale contains double-barreled statements and has "left" and "right" parts to it. Neither of these characterize the Dogmatism scale. Despite these differences, however, the reliabilities of the 40-item Opinionation scale, Forms C and C_E, and the 40-item Dogmatism scale, Form E, are about the same (.75-.80).

4. In general, the opinionated rejec-

---

[11] We did not conduct an item analysis of Form B because a study of the reliabilities for the total scale and the several subscales led us to believe that further development of this scale (see earlier description of its design) would not be worthwhile.

TABLE 6

ITEM ANALYSIS OF THE OPINIONATION SCALE, FORM C FOR THE M.S.U. III GROUP

Items	Total Group (N=186)		Highs (N=47)		Lows (N=47)		Diff.	t^a	Form A^b t	Form A^c t
	M	σ	M	σ	M	σ				
Left Opinionation, Opinionated Rejection										
1. Pro-Roosevelt	5.35	1.96	5.75	1.63	4.68	2.29	1.07	2.58	N.S.	N.S.
2. No race differences	3.92	2.00	4.92	1.70	3.06	1.90	1.86	4.95	3.89	6.41
3. Pro-Truman's Fair Deal	2.31	1.28	2.92	1.46	1.66	.69	1.26	5.32	4.92	4.17
4. Anti-loyalty oaths	3.94	1.89	4.04	1.91	3.70	1.87	.34	N.S.	2.90	2.88
5. Russia not imperialistic	4.13	1.84	4.72	1.71	3.77	1.83	.95	2.57	3.65	2.52
6. Anti-real estate interests	3.20	1.60	4.17	1.42	2.30	1.16	1.87	6.93	—[d]	—
7. Government & education not Red-infiltrated	1.86	1.07	2.15	1.13	1.43	.56	.72	3.85	—	—
8. Anti-universal military training	2.22	1.24	2.89	1.51	1.77	1.09	1.12	4.06	—	—
9. Anti-government support of religion	5.02	2.02	5.72	1.66	3.94	2.14	1.78	4.45	—	—
10. Anti-McCarthy	2.99	1.69	3.62	1.77	2.57	1.56	1.05	3.02	—	—
Left Opinionation, Opinionated Acceptance										
11. Pro-Rosenbergs	2.39	1.55	2.40	1.48	2.77	1.78	−.37	N.S.	—	—
12. Anti-MacArthur	2.61	1.57	3.00	1.70	2.32	1.58	.68	1.99	3.63	2.85
13. Anti-big business	3.93	1.87	4.30	1.87	3.49	2.05	.81	1.98	—	—
14. Pro-social security	2.28	1.40	2.55	1.71	1.87	1.13	.68	2.25	—	—
15. Anti-Congressional investigating committees	2.80	1.52	2.96	1.70	2.36	1.23	.60	1.94	—	—
16. Anti-American Legion	2.08	1.07	2.28	1.06	1.70	.93	.58	2.80	—	—
17. Pro-soak-the-rich taxes	2.51	1.43	3.38	1.49	2.06	1.25	1.32	4.62	—	—
18. Rich getting richer, poor getting poorer	2.32	1.45	2.72	1.57	1.77	.96	.95	3.49	—	—
19. Capitalism causes depressions and wars	2.12	1.16	2.72	1.55	1.72	.72	1.00	3.95	3.53	3.20
20. Anti-Hoover	3.14	1.60	4.04	1.71	2.38	1.20	1.66	5.39	3.69	2.67
Right Opinionation, Opinionated Rejection										
21. Anti-socialized medicine	3.92	2.00	5.23	1.83	2.94	1.64	2.29	6.33	4.05	4.21
22. Pro-Eisenhower	4.04	1.60	4.55	1.48	3.47	1.66	1.08	3.28	—	—
23. Reds yell about civil rights	3.33	1.68	4.38	1.72	2.38	1.52	2.00	5.90	—	—
24. Anti-Labor	4.55	1.77	5.38	1.58	3.21	1.78	2.17	6.18	5.31	6.43
25. Anti-Democratic Party	4.38	1.84	4.79	1.79	4.11	1.89	.68	1.77	—	—
26. Pro-God	4.31	1.96	5.15	1.79	3.68	2.11	1.47	3.61	4.52	4.00
27. Anti-government ownership of utilities	3.23	1.64	4.13	1.61	2.13	1.10	2.00	6.99	—	—
28. U.S. no warmonger	4.75	1.79	5.38	1.45	4.23	1.96	1.15	3.21	4.16	3.07
29. Anti-socialism	3.44	1.70	4.55	1.67	2.36	1.35	2.19	6.93	—	—
30. Anti-Red China	4.14	1.60	4.77	1.39	3.66	1.79	1.11	3.31	4.78	4.41
Right Opinionation, Opinionated Acceptance										
31. U.S. re-arms to stop aggression	4.40	1.89	5.19	1.54	3.94	1.93	1.25	3.42	3.83	1.83
32. Prejudice removable by education	5.74	1.54	6.00	1.29	5.45	1.93	.55	N.S.	3.68	3.73
33. U.S. should send military aid	2.67	1.54	2.98	1.78	2.30	1.42	.68	2.03	—	—
34. Pro-Churchill	4.92	1.29	5.45	1.08	4.66	1.50	.79	2.90	2.75	3.45
35. U.S. re-arms to preserve freedom	4.69	1.71	5.30	1.28	4.04	1.97	1.26	3.64	3.50	3.03
36. Pro-Chiang Kai Shek	3.49	1.64	4.51	1.49	2.75	1.49	1.76	5.66	4.87	4.88
37. U.N. a failure	1.93	1.22	2.36	1.45	1.36	.81	1.00	4.08	—	—
38. Businessman contributes most to society	3.85	1.78	4.45	1.73	2.98	1.69	1.47	4.13	—	—
39. Pro-Franco	4.49	1.66	5.00	1.52	3.70	1.81	1.30	3.74	3.81	2.75
40. Alger Hiss a traitor	4.62	1.57	5.13	1.32	3.75	1.65	1.38	4.44	—	—

---

[a] Employing a one-tailed t test, t's of 1.64 and 2.33 are significant beyond the 5 and 1 per cent levels, respectively, for large samples.

[b] These t's are obtained from an item analysis of the M.S.U. I group (N=202).

[c] These t's are obtained from an item analysis of N.Y. Colleges Group (N=207).

[d] A dash (—) means that the item in question was devised for a subsequent form.

tion items are more discriminating than the opinionated acceptance items.

5. The best items on the Opinionation scale, employing the arbitrary criterion of a t of 5.00 or more for items used in Form C only, and a t of 4.00 or more for items used in both Forms A and C, are:

Item 3. There are two kinds of people who fought Truman's Fair Deal program: the selfish and the stupid.

Item 6. It's mainly those who believe the propaganda put out by the real-estate interests who are against a federal slum clearance program.

Item 21. It's simply incredible that anyone should believe that socialized medicine will actually help solve our health problems.

Item 23. It's the fellow travellers or Reds who keep yelling all the time about Civil Rights.

Item 24. It's the radicals or labor racketeers who yell the loudest about labor's right to strike.

Item 27. It's usually the trouble-makers who talk about government ownership of public utilities.

Item 29. It's mostly the noisy liberals who try to tell us that we will be better off under socialism.

Item 36. This much is certain! The only way to defeat tyranny in China is to support Chiang Kai-shek.

VI. Major Hypotheses

Thus far we have tried to show how our theoretical formulations guided the construction of the Dogmatism and Opinionation scales. The next step is to determine their validity. Several methods are available for determining this: (*a*) validation against external behavioral criteria, (*b*) validation by the method of known groups, (*c*) validation against other standardized scales purporting to measure the same or similar variables, and (*d*) theoretical validation by testing hypotheses stemming from a set of theoretical considerations.

We propose to employ a combination of methods *b, c,* and *d* above. In employing these methods our aim will be twofold: to determine not only the validity of our conceptualization and measurement of the phenomenon of dogmatism, but also the validity of our assumptions regarding the nature of the conceptualizations and measurements employed in the California research.

Each of the three major hypotheses of the study, to be set forth below, addresses itself to this twofold purpose.

Hypothesis A: Dogmatism as a Measure of General Authoritarianism

A major defining characteristic of dogmatism is that it involves adherence to some central set of beliefs in absolute authority. Thus, it should be possible to demonstrate that dogmatism is related to some independent measure of authoritarianism. The F scale has become widely accepted as a measure of authoritarianism. One would therefore expect that scores on the Dogmatism scale would be related, to a considerable extent, to scores on the F scale.

However, the Dogmatism scale should also differ in certain important respects from the F scale. It has been repeatedly shown that the F scale correlates substantially with ethnocentrism (average $r = .73$). If the Dogmatism scale does indeed represent a measure of authoritarianism regardless of specific ideological or attitudinal position, then it should be relatively independent of ethnocentrism and independent of position along the left-to-right continuum. And it should correlate with the F scale independently of these variables.

Hypothesis B: Left, Right, and Total Opinionation in Relation to Dogmatism and the Authoritarian Personality

A major defining characteristic of dogmatism is opinionation. Thus scores on the two scales should be systematically related to each other. However, the manner in which the Dogmatism scale will be related to opinionation should differ markedly from that expected between the F scale and opinionation, or between the E scale and opinionation. In the case of the Dogmatism scale its relation to opinionation should be a function of positive relations to both left and right opinionation; that is, scores on the Dogmatism scale should be positively related to scores on the Left and Right Opinionation subscales as well as to the Total Opinionation scale. In the case of the F and E scales, however, it is expected that they will be related primarily to the Right Opinionation subscale but not at all to the Left Opinionation subscale.

Hypothesis C: Religious and Political Commitment in Relation to Dogmatism and the Authoritarian Personality

It is well known that there exist certain institutionalized, hierarchically organized groups, for example, the Catholic Church and Communist Party, which demand strict adherence to the religious or political doctrines to which they are committed. Whether one becomes a member of such a social organization by birth or by joining it later, it can be broadly assumed that the field forces operating upon the person will be such that they will force him to commit himself in advance, that is, to accept in advance the group's ideology and to reject in advance alternative ideologies. It is beyond the scope of the present research to spell out in detail how such pressures may be exerted. This is a challenging research problem in itself. However this may be achieved, strong group pressures toward commitment are assumed to have manifold effects upon the cognitive organization and development of a person's belief-disbelief system: isolation within and between belief and disbelief systems, narrowing, dedifferentiation of the disbelief system—indeed, all the subvariables which we have assumed to be defining characteristics of dogmatic thinking and believing.

Accordingly, it is hypothesized that the greater the group pressures toward commitment—religious or political, left-of-center or right-of-center—the greater the dogmatism and opinionation.

What can one say about the effects of group pressures toward commitment on authoritarianism and intolerance, as measured by the F and E scales? This will depend, in part, upon whether the commitment is to a left-of-center or right-of-center ideological outlook. In the case of strong commitment to a left-of-center ideological group we hypothesize, in line with our theoretical analysis, relatively high scores on the Dogmatism and Opinionation scales, but at the same time relatively *low* scores on the F and E scales. This is consistent with the assumption that the F and E scales measure primarily right authoritarianism and intolerance. In the case of commitment to a right-of-center ideology we hypothesize not only relatively high scores on the Dogmatism and Opinionation scales but also relatively high scores on the F scale. Whether or not there will also be relatively high scores on the E scale, however, is difficult to predict in advance since it is quite possible for a right-of-center group (e.g., a religious group) to demand from its members strict adherence to its ideological outlook but at the same time advocate equalitarian attitudes toward Jews, Negroes, etc.

VII. HYPOTHESIS A: DOGMATISM AS A MEASURE OF GENERAL AUTHORITARIANISM

A. RESULTS

1. Dogmatism and the F Scale in Relation to Ethnocentrism

In Table 7 are shown the correlations found between dogmatism (D) and authoritarianism (F) for the various groups tested in the United States and England. Also shown are the correlations between each of these two variables and ethnocentrism (E). Finally, in the last column is shown the correlation between dogmatism and authoritarianism with ethnocentrism held constant by the method of partial correlation.

It is seen that the correlations between D and F range from .54 to .77 in the several groups sampled. These are to be compared with the correlations between F and E which range from .56 to .64. The latter findings are consistent with the findings reported by Adorno *et al.* (1), and by Christie and Jahoda (4). The correlations between D and F are about

TABLE 7

INTERCORRELATIONS AMONG DOGMATISM (D), AUTHORITARIANISM (F), AND ETHNOCENTRISM (E)

Group	Correlations				
	N	D,F	F,E[a]	D,E	(D,F)·E
M.S.U. I	202	.67	.56	.36	.61
N.Y. Colleges	207	.58	.63	.46	.42
Jewish subsample	131	.61	.60	.49	.45
M.S.U. II	153	.61	.58	.33	.54
M.S.U. III	186	.54	.64	.31	.47
English Colleges I	137	.57	.58	.39	.46
English Colleges II	80	.62	.62	.32	.57
English Workers	60	.77	.56	.53	.67

[a] There were 21 Jews in the English Colleges I group, 11 Jews in the English Colleges II group, one Jew in the English Worker group. These *S*s were excluded in computing the correlations between F and E, and between D and E (*N* = 116, 69, 59, respectively).

as high as those between E and F, the average *r* for the former being .62, for the latter .60.

That D and F correlate as highly with each other as do E and F would lead one to suspect that D would also correlate with E to about the same extent as F correlates with E. But a glance at the third column of Table 7 shows that in every group tested the correlation between D and E is smaller than that between F and E. This leads us to the last column of Table 7 which shows the correlations between D and F when ethnocentrism is held constant by the method of partial correlation. In the M.S.U. I group, for example, the correlation between D and F drops only from .67 to .61 when E is partialled out. In the various groups the partialled correlations drop to varying degrees, but in every case the decreases are relatively small, and all the partialled correlations remain significant well beyond the one per cent confidence level.

2. *Further Analysis of What the F Scale Measures*

It is well known that each of the F-scale items was validated by determining the extent to which each item discriminated between upper and lower ethnocentrism quarters (1). All the items retained in the final F scale—the one employed in the present research—discriminated significantly between extreme E groups. Thus, it is indeed surprising to find that the F scale correlates to the extent that it does with the Dogmatism scale even when ethnocentrism is partialled out or held constant.

The findings just presented indicate a need to examine more closely what it is that the F scale measures. We did this in the following manner: For the M.S.U. I group (*N* = 202) and the New York Colleges group (*N* = 207) the *S*s were first divided equally into high- and low-E groups according to whether their ethnocentrism scores fell above or below the median. These two groups were then further subdivided into high- and low-dogmatism groups according to whether *S*s scored above or below the median dogmatism score for the total group. In this way four subgroups were obtained; I. Low E, Low D; II. Low E, High D; III. High E, Low D; IV. High E, High D. These four subgroups were then compared on each of the F scale's 29 items.[12]

The full details of this analysis are too long to present here. Instead, we present, in Table 8, a summary of one analysis which, to our mind, is a crucial one.

Table 8 shows that there are 9 F items out of 29 which differentiate significantly (in the M.S.U. I group or the N.Y. group

---

[12] The F scale employed in the present research is Form 40-45. Through an oversight, one item (force to preserve) was omitted.

TABLE 8

F-Scale Item Comparisons Among Groups High and Low in Dogmatism and Ethnocentrism

Group I = Low in ethnocentrism, low in dogmatism
Group II = Low in ethnocentrism, high in dogmatism
Group III = High in ethnocentrism, low in dogmatism
Group IV = High in ethnocentrism, high in dogmatism

F-scale items which differentiate significantly[a] between:	
I and II but not between II and IV	II and IV but not between I and II
1. A person who has bad manners, habits and breeding can hardly expect to get along with decent people.	1. The businessman and the manufacturer are much more important to society than the artist and the professor.
2. Young people sometimes get rebellious ideas, but as they grow up they ought to get over them and settle down.	2. Science has its place, but there are many important things that can never possibly be understood by the human mind.
3. What this country needs most, more than laws and political programs, is a few courageous, tireless, devoted leaders in whom the people can put their faith.	3. No sane, normal, decent person could ever think of hurting a close friend or relative.
4. Nobody ever learned anything really important except through suffering.	4. What the youth need is strict discipline, rugged determination, and the will to work and fight for family and country.
5. When a person has a problem or worry it is best for him not to think about it, but keep busy with more cheerful things.	5. Homosexuals are hardly better than criminals and ought to be severely punished.
6. People can be divided into two distinct classes: the weak and the strong.	6. Some people are born with an urge to jump from high places.
7. Most people don't realize how much our lives are controlled by plots hatched in secret places.	7. Wars and social troubles may someday be ended by an earthquake or flood that will destroy the whole world.
8. Familiarity breeds contempt.	8. Nowadays when so many different kinds of people move around and mix together so much, a person has to protect himself especially carefully against catching an infection or disease from them.
9. The wild sex-life of the old Greeks and Romans was tame compared to some of the goings-on in this country, even in places where people might least expect it.	

[a] Two-tailed *t* tests were employed in this analysis. The statement "F-scale items which differentiate significantly between I and II but not between II and IV" means that such items differentiate significantly beyond the five per cent level between I and II in the M.S.U. I group or the New York College group or both, but do not differentiate significantly between II and IV either in the M.S.U I group or the New York group.

or both) *only* between the Low E, Low D group (I) and the Low E, High D group (II) but not between the Low E, High D group (II) and the High E, High D group (IV). In other words, these nine items differentiate significantly between two groups of *S*s low in E differing in dogmatism, and do not differentiate significantly between two groups of *S*s differing in E but equated on dogmatism.[13]

Other F items—8 out of 29—show exactly the opposite trend. They discriminate significantly, in the M.S.U. I group, or the N.Y. group, or both, *only* between the two groups of *S*s differing in *E* but equated on dogmatism (Group II vs. Group IV), but not between the two groups low in E differing in dogmatism (Group I vs. Group II). Of the remaining 12 F items (29−9−8=12), 10 discriminated significantly not only between Groups I and II but also between II and IV. One item (prewar authorities in Germany should be used in postwar Germany) differentiated significantly only between Groups I and IV. One item (astrology) did not differentiate significantly among any of the four groups either in the M.S.U. I group or the New York Colleges group.

---

[13] Six of these nine items also discriminate significantly between Groups III and IV in the M.S.U. I group, or the New York Colleges group, or both.

The results just discussed shed added light on what the various items on the F scale actually measure. We will discuss these results more fully in the latter part of this section.

3. *Dogmatism and the F Scale in Relation to Liberalism-Conservatism*

We proceed next to present data of another sort bearing on Hypothesis A. In Table 9 are shown the correlations between the Dogmatism scale and two measures of liberalism-conservatism: the P.E.C. scale (1) and the score obtained by subtracting the left-opinionation score from the right-opinionation score. These are to be compared with correlations found between the F scale and these same two measures. Also shown are the correlations found between the two measures of liberalism-conservatism.

One must bear in mind, when interpreting these data, that the P.E.C. scale employed was composed of only 5 items and that the second measure employed (R.O. − L.O.) measures degree of orientation leftward or rightward relatively crudely. Thus, the correlations shown in Table 9 are of interest not so much because they reflect true relationships but rather because they point up differences in the extent to which the D and F scales correlate with liberalism-conservatism.

Attention is drawn first to the last column of Table 9 showing the correlations between the two measures of liberalism-conservatism. While these correlations are by no means high they are in every case higher than the correlations between either of these two measures and D or F. It is on the basis of this latter finding that we feel justified in assuming that the (R.O. − L.O.) measure is, like the better known P.E.C. scale, at least a rough index of liberalism-conservatism.

Table 9 shows that on the whole the Dogmatism scale correlates to a negligible extent with the two measures of liberalism-conservatism. And in every single group, the F scale correlates more highly with liberalism-conservatism than does the D scale.

At the same time it is worth noting

TABLE 9

THE DOGMATISM AND F SCALES IN RELATION TO LIBERALISM-CONSERVATISM

Group	*N*	Correlations between D and		Correlations between F and		Correlation between P.E.C. and (R.O.-L.O.)
		P.E.C.	(R.O.-L.O.)[a]	P.E.C.	(R.O.-L.O.)	
M.S.U. I	202	.13	.13	.22	.28	.38
N.Y. Colleges	207	.11	.04	.43	.37	.58
M.S.U. II	153	.20	.13	.29	.21	.30
M.S.U. III	186	.28	.17	.40	.35	.49
English Colleges I	137	—[b]	.12	—	.36	—
English Colleges II	80	—	−.03	—	.31	—
English Workers	60	—	.11	—	.15	—

[a] Right-opinionation score minus left-opinionation score. Positive differences indicate a rightist orientation; negative differences indicate a leftist orientation.

[b] The P.E.C. scale was not employed in the English research.

that the correlations between dogmatism and conservatism, while quite small, are consistently positive. Several interpretations of this finding are possible and will be considered in Section X.

B. Discussion

We have brought to bear several kinds of evidence relevant to the hypothesis that the Dogmatism scale represents a measure of general authoritarianism, and that the F scale, by contrast, represents a measure of right authoritarianism.

The D scale correlates relatively highly with the F scale—as highly as the correlations between the F and E scales. When E is held constant by the method of partial correlation, the correlations between D and F are reduced but slightly.

The Dogmatism scale appears to be about as good a measure of authoritarianism as the F scale. These two scales correlate as highly with each other as do the F and E scales. At the same time, the Dogmatism scale clearly differs from the F scale in that it is related to ethnocentrism to a smaller extent, and considering the size of the correlations between D and F when E is held constant it can be said that the D scale measures authoritarianism independent of ethnic attitudes.

The dogmatic low. It now becomes possible, perhaps for the first time, to distinguish objectively by means of the Dogmatism scale between persons whose low-ethnocentric beliefs are dogmatic and "pseudo-democratic" from those whose low-ethnocentric beliefs are more "genuine." Thus, low E-scorers differ widely in dogmatism and even though their F-scale scores are relatively low.

The "nondogmatic" high. By the same token, it now becomes empirically possible to speak not only of low-ethnocentric persons as differing in dogmatism (and authoritarianism) but also of high-ethnocentric persons as differing in dogmatism (and authoritarianism). As has been shown, high E-scorers differ markedly in degree of dogmatism, and even though they score relatively highly on the F scale, dogmatism and F scores are systematically related to each other.

These findings suggest the following: (*a*) Highly ethnocentric attitudes, as measured by the E scale, are by no means necessarily a function of an authoritarian personality structure. (*b*) By means of the Dogmatism scale it is possible to differentiate the low-dogmatic, low-authoritarian person who scores high on the E scale from the high-dogmatic, high-authoritarian person who scores high on the E scale. (*c*) Unanalyzed social and personal conditions other than authoritarian personality structure are probably operating to produce high scores on the E scale. What these conditions are is, to our mind, a fruitful question to which future research should address itself. (*d*) Action programs designed to reduce ethnic prejudice (e.g., fair employment practices programs; desegregation in education and housing; etc.) may have differential effects upon persons having anti-ethnic attitudes, depending upon the extent to which such attitudes are imbedded in "authoritarian-dogmatic" personality structure.

Is the F scale a scale? The findings further lead us to ask whether the F scale, considered alone, orders individuals quantitatively along a continuum of authoritarianism. For example, it appears that a highly dogmatic person in the lowest E quarter does not have to obtain as high an F score as a highly dogmatic person in the upper E quarter in order to be equally authoritarian. In other words, our findings suggest that depending upon where one falls on the

E scale continuum, different F scores may represent equal degrees of authoritarianism, and the same F score may represent different degrees of authoritarianism.

The interpretation that the F scale may not represent an ordinal continuum becomes even more compelling when we consider the data for each F item, comparing high and low ethnocentrism groups further subdivided into high and low dogmatic groups. These data make it possible to pin down more precisely what sort of authoritarianism each F item measures. We reasoned that if the F scale is indeed related to ethnocentric attitudes toward Jews, Negroes, etc., then each F item should continue to discriminate between high and low E scores regardless of whether they are equated for dogmatism or any other variable. And certainly none of the F items should be expected to discriminate *only* between two groups of low E scorers, or high E scorers, differing in dogmatism.

Three sets of F items, more or less distinct from each other, emerge from this analysis:

1. Eight of the 29 F items discriminate between high and low E groups, when these groups are equated for dogmatism (Group II vs. IV) but not between two groups of low E scorers differing in dogmatism (Group I vs. II). Thus, it would appear that it is these items which are "bearing the burden" in giving the F scale its characteristic items which tap primarily association with ethnocentrism. The content of these items has to do with the glorification of the business man and the manufacturer, the belief that beyond science is the supernatural, an inability to express hostility toward close relations, a belief in strict discipline for youth, a punitive attitude toward homosexuals, the belief that some people are born with an urge to jump from high places, the belief that some day the world will be overcome by catastrophe, and a fear of infection and disease.

2. Nine F items dicriminate between two groups of low E scorers, high and low dogmatism (Group I vs. II) but not between high and low E groups high in dogmatism (Group II vs. IV). This finding is a surprising one in view of the fact that *all* the F items were originally validated against extreme groups on the E scale. Again the content of these items is worth noting; they have to do with bad manners and breeding, the belief that youth should outgrow its rebelliousness, the need for tireless, devoted leaders, learning through suffering, keeping busy to avoid self-scrutiny, the division of people into weak and strong, paranoid ideas about secret plots, familiarity leading to contempt, and the projection of sexual impulses onto others.

3. Ten of the remaining 12 F items discriminate significantly between high and low-ethnocentric groups equated for dogmatism, and between high- and low-dogmatic groups equated for ethnocentrism.

From these findings there emerges a clearer understanding of what the F scale measures. Approximately one-third of the items (eight items) seem to measure right authoritarianism, another third general authoritarianism (nine items) and the remaining third something in between right and general authoritarianism (10 items). The net effect seems to be a scale "loaded" in such a way that it taps primarily right authoritarianism. Thus, the F scale is seen to be useful as a measure of individual differences in authoritarian personality structure *primarily among persons to the right-of-center.* It becomes increasingly insensitive as a measuring device the more one's atten-

TABLE 10

LEFT, RIGHT, AND TOTAL OPINIONATION: CORRELATION WITH OTHER VARIABLES

Group	N	Correlation Between											
		Dogmatism and—			F scale and—			Ethnocentrism[a] and—			P.E.C. Scale[b] and—		
		Left Opin.	Right Opin.	Total Opin.	Left Opin.	Right Opin.	Total Opin.	Left Opin.	Right Opin.	Total Opin.	Left Opin.	Right Opin.	Total Opin.
M.S.U. I	202	.22	.40	.51	.11	.53	.54	−.08	.50	.37	−.20	.41	.19
N.Y. Colleges	207	.20	.23	.43	−.08	.56	.50	−.22	.62	.43	−.39	.60	.25
M.S.U. II	153	.20	.31	.34	.12	.46	.45	.02	.44	.41	−.10	.26	.21
M.S.U. III	186	.24	.42	.47	.12	.55	.49	.01	.50	.39	−.18	.48	.26
English Colleges I	137	.17	.38	.61	−.27	.64	.40	−.21	.58	.38	—	—	—
English Colleges II	80	.25	.25	.55	−.06	.56	.46	−.22	.49	.20	—	—	—
English Workers	60	.22	.43	.63	.18	.47	.63	−.18	.59	.50	—	—	—

[a] The *N*'s for the English Colleges I, II, and English Worker samples are 116, 69, and 59 respectively, since Jewish *S*s were omitted.

[b] Not given in England.

tion shifts away from the study of right authoritarianism to the study of authoritarianism along other points on the left-to-right continuum.

Additional evidence, regarding the differential properties of the Dogmatism and F scales is found in their differential correlations with liberalism-conservatism, shown in Table 10. The Dogmatism scale correlates negligibly with conservatism and in every case these correlations are smaller than those between the F scale and conservatism. These findings indicate also that dogmatism is to be found with approximately equal frequency along all positions of the left-to-right continuum.

Nevertheless, it should not be overlooked that the correlations between dogmatism and conservatism, while small, are consistently positive. This finding is probably related to the additional finding of positive correlations between D and E, and is further related to other findings to be presented shortly in connection with Hypothesis B. We will therefore delay further consideration of the possible theoretical significance of these findings until Section X.

VIII. HYPOTHESIS B: LEFT, RIGHT, AND TOTAL OPINIONATION IN RELATION TO DOGMATISM AND THE AUTHORITARIAN PERSONALITY

A. RESULTS

As already stated the Opinionation scale is essentially a measure of one aspect of dogmatism, namely, intolerance of those who disagree with, and qualified tolerance of those who agree with, one's views. The Left- and Right-Opinionation subscales are measures of the extent to which one is opinionated in a left or right direction, respectively. But considered as separate scores these subscales do not tell us much about individual differences in opinionation. It is the total opinionation score—the sum of left and right opinionation—which provides this information.

The California E scale also represents a measure of intolerance. We have, then, two measures of intolerance. The Opinionation scale defines intolerance in terms of acceptance or rejection of others who agree or disagree with one's views; the E scale defines intolerance in terms of acceptance or rejection of ethnic groups.

Hypothesis B is to the effect that dogmatism should be positively related to total opinionation, and to left and right opinionation as well. The F and E scales, however, should be related only to right opinionation and not at all, or even inversely, with left opinionation. That is, the F and E scales should be sensitive as predictors of individual differences in right opinionation but relatively insensitive as predictors of individual differences in left opinionation.

The results bearing on this issue are shown in Table 10. Also shown are the correlations between the P.E.C. scale and left, right, and total opinionation.

The following points are noted:

1. To be kept in mind is that the total Opinionation scale is composed of 40 items, 20 of which are left- and 20 right-opinionation items. As has already been shown in Table 4, left and right opinionation are either negligibly related to each other in some groups, or related negatively to a considerable extent in other groups.

2. The total Opinionation scale correlates with the D scale to about the same extent as with the F scale. In the former case the range of *r*'s is from .34 to .63; in the latter case the range is from .40 to .63. The total Opinionation scale correlates to a somewhat lesser extent with the E scale, the range of *r*'s being from .20 to .50. And it correlates to an even smaller, but nevertheless positive extent with P.E.C., the range being from .19 to .26.

3. The most important finding bearing on Hypothesis B is that in all groups dogmatism correlates positively with both left and right opinionation, despite the fact that left and right opinionation are negligibly or, more often, negatively related to each other. In contrast, both the F and E scales correlate relatively highly with right opinionation but negligibly or negatively with left opinionation. Furthermore, the pattern of these latter correlations is very similar to that between P.E.C. and left and right opinionation, except that P.E.C. consistently correlates more negatively with left opinionation (but not much more), and consistently correlates to a lesser degree with total opinionation.

4. In every group the Dogmatism scale correlates more highly with total opinionation than with left or right opinionation. This is not the case with the F, E, and P.E.C. scales. In practically all instances these variables correlate less with total opinionation than with right opinionation, due to the "suppressing" effect of left opinionation.

5. To be noted, finally, is that while the Dogmatism scale correlates positively with both left and right opinionation it correlates consistently more highly with right than with left opinionation.

B. Discussion

In accord with Hypothesis B, dogmatism scores are found to be positively related to total opinionation scores. Also in accord with Hypothesis B, dogmatism is in every case positively related to both left and right opinionation. The latter finding is all the more interesting in view of the fact that left and right opinionation are negatively correlated with each other to a substantial degree in the New York Colleges group and in the three English groups.

The total opinionation scores are correlated with the F scores to about the same extent as with the D scores, and to a somewhat lesser extent with the E scores. If we had only the total opinionation scores to go by, one could easily be misled in interpreting these findings. It is, however, clear from the left and right opinionation correlations that the origins of the correlations with total opinionation are different. They derive clearly from the fact both the F and E scales are relatively highly correlated with right opinionation, and are uncorrelated or even negatively correlated with left opinionation. These findings suggest, parallel to the distinction drawn between the Dogmatism and F scales, that the Opinionation scale represents primarily a measure of general intolerance, while the E scale represents primarily a measure of right intolerance.

Nevertheless, it cannot be ignored that dogmatism correlates in consistent fashion somewhat more highly with right than with left opinionation. This finding is in accord with the findings presented earlier on the small positive correlations between dogmatism and liberalism-conservatism. As already stated, these findings will be considered in Section X.

One final point needs to be made here: from the correlations between total opinionation and E (Table 10) it is evident that while the two forms of intolerance tend to go together, the relationship is far from perfect; that is, a person can score low in ethnocentrism but be relatively high in opinionation, and vice versa. We thus have in the Opinionation scale an instrument, in addition to the Dogmatism scale, by means of which it is possible to isolate for further study relatively dogmatic low E scorers,[14] and by the same logic it is also possible to isolate for further study relatively undogmatic high E scorers.

---

[14] In Hypothesis C we try to come to closer grips with this issue.

IX. HYPOTHESIS C: RELIGIOUS AND POLITICAL COMMITMENT IN RELATION TO DOGMATISM AND THE AUTHORITARIAN PERSONALITY

A. RESULTS

Hypothesis C is to the effect that persons having strong commitment to religious or political groups will manifest more dogmatism and opinionation than persons having lesser commitments. In the case of strong commitment to right-of-center groups, we expect relatively high scores on dogmatism and opinionation, and at the same time relatively high scores on the F scale and, possibly (but not necessarily), on the E scale. In the case of commitment to left-of-center groups, we also expect relatively high scores on dogmatism and opinionation, but relatively *low scores* on F and E.

1. On Religious Commitment

Let us consider first the breakdown by religion of the M.S.U. I and the New York Colleges groups. We first assumed that the Catholic Church represents an ideological position conservative in nature or, at least, generally to the right-of-center. We assumed further that the Catholic Church is generally speaking relatively effective, by virtue of its social organization, in demanding strict adherence to its dogma. In Table 11 are shown the mean scores on the Opinionation, Dogmatism, F and E scales for the M.S.U. I group broken down into Catholic, Protestant, and None (no religion) subgroups.[15] In Table 12 are shown comparable data for the New York Colleges group, broken down into Catholic, Protestant, Jewish, and "None" subgroups.

Consider first the data for the M.S.U. I subgroups, shown in Table 11. It is seen that on all four of the major variables—opinionation, dogmatism, F, and E—the

---

[15] It will be recalled that Jews (and Negroes) were excluded from this study.

TABLE 11

Comparisons Among Various Religious Groups in the M.S.U. I Group on Opinionation, Dogmatism, F and E

Group	N	Left Opinionation		Right Opinionation		Total Opinionation		Dogmatism		F Scale		Ethnocentrism	
		M	σ	M	σ	M	σ	M	σ	M	σ	M	σ
1. Catholics	42	58.43	11.82	94.71	12.89	153.14	12.21	191.10	27.07	109.79	20.75	34.31	11.14
2. Protestants	145	58.47	12.02	86.41	12.81	144.88	17.04	180.10	25.42	99.24	21.80	29.99	11.03
3. None	15	67.73	9.30	76.67	11.80	144.40	15.29	174.60	21.10	91.60	21.39	24.27	7.54
Comparisons	df	t[a]	p	t	p	t	p	t	p	t	p	t	p
1 vs. 2	185	—	N.S.	3.64	.01	3.47	.01	2.33	.01	2.84	.01	2.19	.05
1 vs. 3	55	3.00	.01	4.82	.01	1.94	.05	2.34	.01	2.77	.01	3.77	.01
2 vs. 3	158	3.46	.01	2.92	.01	—	N.S.	—	N.S.	—	N.S.	2.58	.01

[a] All t tests are one-tailed.

Catholic group is highest, the Protestant[16] group is intermediate and the "None" group is lowest.[17] On all four of these variables the Catholic group is significantly higher than the Protestant group, and significantly higher than the "None" group. The Protestant group is significantly higher than the "None" group only on the E variable.

Of further interest are the mean scores on left and right opinionation. It is clear that all three groups are more right- than left-oriented, as is indicated by the fact that the mean right-opinionation scores are higher than the mean left-opinionation scores. However, the "None" group is significantly more left-opinionated than the Catholic and Protestant groups, the latter two groups not differing from each other in this respect. On right opinionation, the Catholic group scores highest, the Protestant group next highest, the "None" group lowest. Each of these means is significantly different from all others.

Despite these differences on left and right opinionation, the mean total opinionation scores for the Protestant and "None" groups are almost identical and both of these groups score significantly lower on this variable than does the Catholic group.

The findings for the M.S.U. I group are clearly in accord with Hypothesis C. However, the comparable findings for the several religious subgroups of the New York Colleges group seem less conclusive, as shown in Table 12.

1. The Catholic, Protestant, and Jewish groups are more right-opinionated than left-opinionated. The "None" group is considerably more left-opinionated than right-opinionated. These findings suggest that the "None" group is left-oriented. All the differences in means on left and right opinionation among the several groups, with the exception of that between the Catholics and Protestants, are significant beyond the 5 or 1 per cent level.

2. The groups do not differ signifi-

---

[16] Of the total group of 145 Protestants, 75 wrote in "Protestant" on the face sheet in response to the question on religion. The remainder wrote in Episcopalian or Presbyterian ($N = 22$), Methodist ($N = 16$), Congregational ($N = 6$), Baptist ($N = 9$), Lutheran ($N = 10$), or some other denomination ($N = 7$). Means and sigmas for D, O, F, and E were computed separately for the Protestant denominations. None of the differences was statistically significant. The mean D, O, F, and E scores for the Catholic group were found to be higher than for any one of these denominations considered separately.

[17] Socioeconomic status, as measured by a question on parents' income, was found to be unrelated to these variables. Thus, the differences found among the several religious groups cannot be attributed to differences in social status.

TABLE 12

COMPARISONS AMONG VARIOUS RELIGIOUS GROUPS IN THE NEW YORK COLLEGES GROUP ON OPINIONATION, DOGMATISM, F AND E

Group	*N*	Left Opinionation		Right Opinionation		Total Opinionation		Dogmatism		F Scale		Ethnocentrism	
		M	σ	M	σ	M	σ	M	σ	M	σ	M	σ
1. Catholics	46	58.26	14.21	86.37	19.88	144.64	15.86	147.38	29.97	105.24	27.82	26.52	9.96
2. Protestants	24	56.71	15.98	88.84	18.95	145.55	22.41	138.34	32.21	95.80	23.21	25.00	8.16
3. Jews	131	70.18	15.82	76.41	16.00	146.59	17.97	139.53	24.48	94.73	22.23	20.32	7.47
4. None	6	94.35	16.78	61.01	18.46	155.36	14.85	147.20	30.58	93.69	20.21	18.67	5.73
Comparisons	*df*	*t*[b]	*p*	*t*	*p*	*t*	*p*	*t*	*p*	*t*	*p*	*t*	*p*
1 vs. 2	68	—	N.S.	—	N.S.	—	N.S.	—	N.S.	—	N.S.	—	N.S.
1 vs. 3	175	4.71	.01	3.04	.01	—	N.S.	1.58	.10	2.29	.05	3.83	.01
1 vs. 4	50	4.63	.01	2.89	.01	—	N.S.	—	N.S.	—	N.S.	2.62[a]	.05
2 vs. 3	153	3.73	.01	2.97	.01	—	N.S.	—	N.S.	—	N.S.	2.57	.01
2 vs. 4	28	4.58	.01	3.04	.01	—	N.S.	—	N.S.	—	N.S.	2.04	.05
3 vs. 4	135	3.17	.01	1.84	.05	—	N.S.	—	N.S.	—	N.S.	—	N.S.

[a] Corrected by the Cochran-Cox approximation for significance of differences in means when the variances differ significantly (6).
[b] All *t* tests are one-tailed.

cantly from each other on total opinionation. However, the mean opinionation score for the "None" group ($N = 6$) is seen to be higher than those for the remaining groups.

3. The Catholic and "None" groups score equally high on dogmatism, and both groups are higher on this variable than the Protestant and Jewish groups. The only difference which approaches significance (10 per cent level) is that between the Catholic and Jewish groups, the former being more dogmatic.

4. As was found with the M.S.U. I group (Table 11), the Catholic group scores highest on the F and E scales, followed in order by the Protestant, Jewish, and "None" groups. For the F scale, the difference between the Catholic and Jewish groups reaches the 5 per cent confidence level. For the E scale, Catholic and Protestant groups do not differ from each other, but both groups are significantly higher on E than the Jewish[18] and "None" groups.

5. It is interesting to note too that the standard deviations on E decrease along with decreases in mean E scores, the Catholic group having the highest standard deviation, the "None" group the lowest. The difference between these two groups is statistically significant.

6. Quite aside from the question of statistical significance, the over-all trend shown in Table 12 is that the Catholic group is relatively high on dogmatism, F, and E, while the "None" group is relatively high on opinionation and dogmatism, but relatively low on F and E.

2. *On Political Commitment*

To test Hypothesis C further we broke down the English Colleges I sample into the following political groups: Conservatives, Liberals, Attlee Laborites, Bevan Laborites, and Communists. We assumed that of these political groups the Communist Party group was left-of-center and at the same time most effective in demanding adherence to its political ideology. Like the Catholic group it is expected, in accord with Hypothesis C, that the Communist group would also score relatively high in dogmatism and opinionation. Unlike the Catholic group however the Communist group is expected to score relatively low on the F and E scales.

---

[18] It will be recalled that the 10-item E scale employed in the New York study did not contain any items about Jews.

TABLE 13

Comparisons Among Various Political Groups in England on Opinionation, Dogmatism, F and E

Group	*N*	Left Opinionation		Right Opinionation		Total Opinionation		Dogmatism		F Scale		Ethnocentrism[a]	
		M	σ	M	σ	M	σ	M	σ	M	σ	M	σ
1. Conservatives	54	56.15	13.42	83.04	18.77	139.20	17.55	258.76	49.70	115.51	25.03	29.94	8.98
2. Liberals	22	68.67	11.54	68.18	13.94	136.85	13.38	242.91	29.21	98.40	13.98	24.75	7.89
3. Attleeites	27	76.78	12.64	59.15	11.76	135.94	18.54	252.69	36.55	101.79	21.44	22.74	9.27
4. Bevanites	19	96.52	19.42	44.52	10.00	141.05	17.34	249.78	37.93	90.37	25.33	23.50	9.42
5. Communists	13[b]	107.86	18.12	46.62	7.34	154.48	18.35	261.62	32.58	82.93	20.26	16.50	4.23
Comparisons	*df*	*t*[c]	*p*	*t*	*p*	*t*	*p*	*t*	*p*	*t*	*p*	*t*	*p*
1 vs. 5	65	9.32	.01	10.90	.01[d]	2.66	.01	—	N.S.	4.80	.01	7.11	.01[d]
2 vs. 5	33	6.75	.01[d]	5.81	.01[d]	2.91	.01	1.65	.10	2.34	.05	3.33	.01[d]
3 vs. 5	38	5.37	.01	4.00	.01[d]	2.89	.01	—	N.S.	2.62	.01	2.57	.05[d]
4 vs. 5	30	—	N.S.	—	N.S.	2.01	.05	—	N.S.	—	N.S.	2.36	.05[d]

[a] The *N* on which the mean E scores are based are 51, 16, 23, 14 and 10 for groups 1, 2, 3, 4, and 5 respectively. Jewish *S*s were excluded from these computations.

[b] The *N* for the English Colleges I is 137. Seven of these gave "other" political affiliations (e.g., Anarchist, Trotskyite, ex-Communist, Celtic Alliance). Five of the 13 Communists were from English Colleges II, and the full 66-item Dogmatism scale was scored (Form D) for these five *S*s for purposes of the present analysis.

[c] All *t* tests are one-tailed.

[d] Corrected by the Cochran-Cox approximation for significance of differences in means when the variances differ significantly (6).

Table 13 shows the mean scores, for the five political groups for the variables under consideration. A study of these data reveals the following:

1. The Communists are found to score highest of all the groups on total opinionation and dogmatism, but lowest on F and E. On total opinionation the Communists are significantly higher than each of the other groups (5 or 1 per cent level). On dogmatism, the Communists are higher than the Liberals, the difference being significant between the 5 and 10 per cent level. On F and E, however, the Communists are lower than each of the remaining political groups and, in general, significantly so (5 or 1 per cent level).

2. The Communists are higher on left opinionation and lower on right opinionation than each of the remaining groups. All differences are significant except those between the Bevanites and Communists. Conversely, the Conservatives show exactly the opposite trend. They are lowest on left opinionation and highest on right opinionation. The five political groups, arranged in order from right to left, follow in clear-cut ascending order on left opinionation and descending order on right opinionation.

3. Despite these findings for left and right opinionation it is interesting to note that on total opinionation the five groups do not follow any particular ascending or descending order in relation to the right-left continuum. When left and right opinionation are combined to yield a total opinionation score the Communists are seen to be most opinionated, followed in order by the Bevanites, Conservatives, Liberals and Attleeites.

4. As is the case with total opinionation, the mean dogmatism scores, too, do not seem to follow any discernible pattern which can be related to leftness-rightness. The Communists are the most dogmatic and they are followed in order by the Conservatives, Attleeites, Bevanites, and Liberals.

5. In contrast to points 3 and 4 above, a fairly clear-cut relationship is seen between leftness-rightness and scores on the F and E scales. The Conservatives score highest on F and E and these means generally decrease, proceeding from right to left.

6. Finally, attention is drawn to the fact that while the Communists are highest on dogmatism, their mean score

(261.62) is only slightly higher than that for the Conservatives (258.76). The difference in means between the Conservatives and Liberals is significant at the 5 per cent level ($t = 1.70$).

B. Discussion

It is clear that the Catholic and Communist groups diverge sharply from each other with respect to their performance on the California F and E scales. The Catholic group is highest of all the religious groups on both F and E; the Communist group is lowest of all political groups on both F and E.[19]

If we had only the F and E scale means to go on, we would have to conclude that the Catholic groups are relatively high in authoritarianism and prejudice, while the Communist group is relatively low. We would also be forced to conclude that the Conservatives are most and the Communists are least authoritarian and intolerant.

Clearly in contradiction with the preceding are the findings with respect to dogmatism and opinionation. Even though the Catholic and Communist groups score at opposite poles on F and E, they score in the same direction, in general, on dogmatism and opinionation. And while the English Conservatives and Communists are also at opposite poles on F and E, no relation is apparent between leftness-rightness and dogmatism or opinionation. Indeed, both Communists and Conservatives score relatively high on dogmatism.

These findings indicate once again that the F and E scales seem to be systematically related to leftness-rightness and that the Dogmatism and Opinionation scales are relatively more independent of this dimension.

Concerning the Catholic groups. While the New York Catholic group does not differ from the remaining New York groups on opinionation, it is clearly highest on the F and E scales, and it is also highest on the Dogmatism scale, though the difference is significant only between the 5 and 10 per cent levels. Thus, the trend of results for these data is on the whole in the same direction as the M.S.U. I data, but is less conclusive as far as Hypothesis C is concerned. At this point, we can only speculate about the reason why the New York results are not as clear-cut as those found for the M.S.U. I group. One reason which comes readily to mind is that students attending New York colleges are less representative of college students in general. It is generally believed that New York college students are more liberal than college students found elsewhere. This is supported by our data which show, for example, that the New York Catholics are lower than the M.S.U. I Catholics on right opinionation, total opinionation, F, and E.[20]

It is worth mentioning too that while the Catholics are generally higher on the major variables under consideration, the ranges of individual differences for Catholics in both groups are about as great as they are for the other religious groups. Thus, it is seen that other unanalyzed variables, in addition to group pressures toward commitment, are operating to produce the results found.

---

[19] We hesitate to make formal, direct comparisons between Communist and Catholic groups because the former was obtained in England, the latter in America. Nevertheless, it is worth pointing out that a direct comparison of the mean F and E scores for Catholic and Communist groups, shown in Tables 11, 12, and 13, reveal marked differences between them.

[20] Six of the ten E-scale items employed in the two studies were identical (non-Jewish items); four were not. Thus, the mean E scores are only roughly comparable. The Dogmatism scale employed in the two groups was different in length and wording and thus direct comparisons are not justified.

Concerning the Communist group. The English Communists are lower than the Socialists, Liberals, and Conservatives on the F[21] and E scales. These findings are as predicted. It is especially interesting to note that the variability of E scores and right opinionation scores for the Communists are considerably lower than those found for the remaining groups. These are consistent with what is known about the content of Communist ideology.

However, while the Communists are generally the most tolerant toward ethnic and racial groups, as determined by scores on the E scale, they are most intolerant toward those they disagree with, as determined by scores on the Opinionation scale. Thus, we observe a discrepancy between ideological structure and content. If we relied solely on the E scale as a measure of prejudice we could conclude only that the Communists are the most tolerant of all the groups tested. But with the information gained from the Opinionation scale a clearer, more complete, and somewhat different picture emerges.

It may strike one as strange, in view of the positive correlations reported earlier among the four major variables under consideration, that the Communists should score highest on dogmatism and opinionation, and lowest on F and E. These findings become more understandable when it is pointed out that such opposing trends are quite possible even though all the correlations are positive. Even more convincing, perhaps, is the fact that for the Communists *all* intercorrelations among these variables are *positive*. The product-moment correlations between D and O ($N = 13$), D and F ($N = 13$), D and E ($N = 10$), and O and F ($N = 13$), O and E ($N = 10$), and F and E ($N = 10$) are .66, .56, .54, .33, .25, and .46, respectively. The correlations involving E are especially interesting in view of the extremely low E mean and low variability for the Communists.

A final comment about the Communists is the following: we have already had occasion to mention that on all variables the range of individual differences for the Catholic groups obtained in America is generally as great as it is for the other religious groups. A similar point may well be made for the Communist group obtained in England. Except for the E scale and the Right-Opinionation subscale where the variabilities are very low (probably due to the fact that high scores on these variables are incompatible with the content of Marxist ideology), the variability of the Communists on dogmatism, left opinionation, total opinionation and F are about as great as those found for the other political groups. Thus, it is necessary to emphasize that the differences we have considered are group differences and that other factors are also at work in producing differences among individual Communists.

X. General Discussion

Guided by our theoretical analysis of the structure, formal content, and function of dogmatic thinking and believing we have constructed two measuring devices—the Dogmatism and Opinionation scales. The former is by far the more broadly conceived of the two. It represents, by definition, a measure of openness-closedness of belief-disbelief systems, general authoritarianism, and general intolerance. The Opinionation scale, on the other hand, was specifically designed to measure but one major facet of our conceptual definition of dogmatism, namely, degree of tolerance or intolerance toward those who agree or disagree with one's belief-disbelief system.

---

[21] Eysenck reports that the mean F scores for the Communists in England are higher than for other groups (7). We are dubious of this as well as other findings by Eysenck. For a justification of this position see Rokeach and Hanley (22) and Christie (5).

To what extent do these scales measure what they purport to measure? It was to answer this question that the three hypotheses of this study were set up. Actually, each of the three hypotheses addresses itself to different facets of our conceptual representation of dogmatism. Also, each of the three hypotheses addresses itself to the question of determining to what extent the present formulations are compatible with the theory, the measuring instruments, and the findings by the authors of *The Authoritarian Personality* (1). Thus, in testing all three hypotheses the major scales of the California researchers are pitted against those employed in the present research. In this way we hoped to determine empirically the respective properties of the two sets of scales and with this the scope and validity of the theoretical formulations underlying both sets.

The evidence presented comes from a number of samples tested in three more or less distinct social climates: the traditionally conservative Midwest, the more liberal atmosphere prevailing in colleges in Metropolitan New York, and the yet more liberal atmosphere prevailing in England. We have been struck by the consistency of the findings despite differences in social climates, and despite fluctuations in the reliability of the measuring instruments employed.

The results obtained to test Hypothesis A show that the Dogmatism scale measures authoritarianism to the same extent as does the F scale. But there is a difference: The Dogmatism scale seems to measure authoritarianism all along and more or less independent of the left-to-right continuum; the F scale seems to measure essentially what its original name implies—fascistic authoritarianism—and not what is implied by its more generic name—the Authoritarianism scale (24).

The data bearing on Hypothesis B yield information on another facet of dogmatism. The Dogmatism scale correlates positively with the Opinionation scale, and positively with the Left- and Right-Opinionation subscales. Regardless of whether the intolerance is directed to one's political right or left, it is positively associated with dogmatism. This is what is demanded on theoretical grounds. At the same time, the F and E scales also correlate positively with total opinionation. But it is clear that these results are obtained by virtue of the fact that the F and E scales correlate only with right opinionation. Persons opinionated to varying degrees in a leftward direction are generally indistinguishable from each other on authoritarianism, as measured by the F scale. In some groups, in fact, it would even appear that the more the opinionation is in a leftward direction the less the F authoritarianism.

In testing Hypothesis C we employed for the first time external criteria—known religious and political groups. We assumed that the field forces operating on individual Catholics and Communists differ in at least one important respect from those operating on adherents of the other religious and political groups, that is, in the pressures exerted by the group toward commitment. Thus, divergent as these two groups are in their political and religious outlook, members of both groups should manifest generally greater dogmatism and opinionation. However, it was also anticipated, in line with our analysis, that Catholic and Communist groups would score at opposite poles on the F and E scales. Thus, the testing of Hypothesis C involves considerably more than an attempt to ascertain the effects of commitment on dogmatism and opinionation. It further makes it possible to test in a crucial way the contrasting properties of the Dogmatism and Opinionation scales, as against the F and E scales.

As expected, the Catholic groups turn out to be highest of all groups on the F and E scales; the Communists (and the New York "Nones") lowest on the F and E scales. Nevertheless, both groups generally score high on the Dogmatism and Opionation scales.

That the group differences turn out as they do lends support to the hypothesis that differences in social organization and the accompanying differences in amount of pressure exerted upon group members play an important role in determining dogmatism and opinionation. That Communists and Conservatives are both relatively high in dogmatism and that there are, furthermore, wide individual differences within *all* of the groups considered indicates that other factors—personal and social—also play a role. All we can conclude from the present data is that group pressures toward commitment seem to be *one* determinant of dogmatism and opinionation. There are, no doubt, other determinants, both personal and social, which will have to be studied further. It seems clear, too, from the within- and between-group data we have presented that personal and social determinants can not readily be subsumed under each other.

We turn now to consider some general theoretical and methodological issues which merit further discussion.

Dogmatism as a cognitive or emotional variable. It may be asked whether dogmatism, as employed in the present research, is a cognitive or emotional variable. Or put another way, have we not confused the cognitive and the emotional in inserting items involving structure, formal content, and function into the Dogmatism scale? The answer to this question, it seems to us, depends upon one's theoretical outlook. For our part, we regard the distinction between what is cognitive and what is emotional as a convenient one but by no means a necessary one. It is possible to conceive of all emotional states as having their representation in the cognitive belief-disbelief system (and vice versa). Thus, *authoritarianism* as a personality state can also be conceived as referring to *cognitive beliefs* in and about authority; a *fear* of aloneness and isolation can be represented as a *cognitive belief* that "Man on his own is a helpless and miserable creature" (Item 43 of the Dogmatism scale); *anxiety* about the future can be transformed into the *cognitive belief* that "It is only natural for a person to be fearful of the future" (Item 47). In precisely the same way such emotional states as self-hate, a need for self-aggrandizement, and a paranoid outlook on life can all be seen as having their representation in the cognitive structure in the form of some specific belief or disbelief.

Seen in this way the distinction between what is cognitive and what is emotional is all but erased. The strategy of reducing the latter to the former makes our concept of the belief-disbelief system a far more dynamic one than most present-day conceptions of cognitive structure. And in closing the gap between what is emotional and what is cognitive a way is paved for the more objective study of complex, value-laden social phenomena which can be more easily attacked by examining a person's cognitions[22] than his emotions. For all the things a person *feels* (and *wants)* must surely be represented by what he *believes* and *knows* about the world he lives in. In line with these ideas, we are presently entertaining the notion that a full description of a person's belief-disbelief system is also a full description of his personality. Thus, the task for future research is seen to be the study of the conditions which lead to the formation of belief-disbelief systems, their growth, and their modification.

The relation between ideological content and ideological structure. The weight of the evidence presented is that dogmatism, with all that it involves, cuts across specific ideological orientations along the left-to-right continuum. From this is might appear that dogmatism, whether found on the right, in the center, or on the left is genotypically the same, differing only in superficialities. However, several lines of evidence seem to suggest that dogmatism has a somewhat greater affinity to right-oriented than to left-oriented ideologies. The correlations between dogmatism and conservatism (Table 9), and between opinionation and conservatism (Table 10), while small or negligible, are consistently positive; the correlations between dogmatism and right opinionation are consistently larger than those with left opinionation (Table 10). The correlations between dogmatism and ethnocentrism (Table 7) and between opinionation and ethnocentrism (Table 10) are even higher, ranging from .31 to .53 in the former case, and from .20 to .50 in the latter case.

One possible explanation is the following: The results found may be a function of the fact that the samples tested in the United States and England are predominantly right-oriented rather than left-oriented. There would thus be greater *heterogeneity* in the degree to which there is acceptance of right-oriented ideology and greater *homogeneity* in the degree to which there is rejection of left-oriented ideology.

The available evidence seems to rule out this possibility. All four of the American samples and the English-worker sample do indeed show higher mean scores and greater variability on right opinionation than on left opinionation (Table 5). However, this is clearly not the case for the two English college groups, which are more left- than right-opinionated and which, furthermore, are more variable on left opinionation than on right opinionation. Despite such differences all sets of data show that both dogmatism and opinionation are positively correlated, to a small but consistent extent, with right-oriented beliefs or ideology.

---

[22] It must be pointed out, however, that some of the dogmatism items are not purely cognitive. In Form E, the final one, eight out of 40 items (Items 29, 39, 46, 49, 50, 51, 57, 59) have to do with feelings rather than with beliefs as such. However, it will be noted that most of these items nevertheless make some reference to a person's cognitive behavior, e.g., repeating oneself, or not listening in a discussion, how one feels toward people one disagrees with, etc.

A second possibility comes to mind. Perhaps political conservatism and especially ethnocentrism is intrinsically more antidemocratic or authoritarian or antihumanitarian in ideological *content* than liberal and antiethnocentric ideology. Hence dogmatic ideological structures may be psychologically more reconcilable, may be more easily "attachable" to ideologies antidemocratic rather than democratic in content. This possibility becomes all the more compelling when we consider the extreme form which conservatism has taken—fascism. Dogmatic ideological structure and authoritarian, antihumanitarian content complement each other and thus do not engender psychological conflict. Whatever the motivations may be—social and personal—which may lead one to embrace antihumanitarian ideologies may also lead one to develop what we have called a dogmatic belief-disbelief system.

On the other hand it is probably more conflictful psychologically to embrace a liberal, humanitarian cause in a dogmatic (closed, authoritarian, intolerant) manner. For here there is a sharp discrepancy between content and structure. The writer has observed, for example, that it is precisely this discrepancy between ideological content and structure which has often led to disillusionment with causes on the left. Come to think of it, we hear much more frequently of disillusionment with left ideology than with right ideology.

The preceding suggests a major direction which future research in this area should take, namely, a systematic study not only of similarities among persons adhering to various totalitarian ideologies on the right, in the center, and on the left, but also a systematic study of differences among adherents to totalitarian ideologies depending upon whether such ideologies are humanitarian or antihumanitarian in content.

Dogmatism and extremeness of beliefs. To social psychologist and layman alike, dogmatism is often seen to be synonymous with, or a function of, extremeness or intensity of political or religious views (10, 26). We are led to reject this view for two reasons. First, we have no evidence that extremeness as such is associated with dogmatism. As has already been indicated dogmatism is at best only slightly related to liberalism-conservatism. Second, the contention that dogmatism may be a function solely of extremeness of beliefs commits one in advance to the theoretical position that extremely held beliefs are necessarily indicative of *dogmatic conviction,* and rules out in advance the possibility that some extremely held beliefs may also be a function of *intellectual conviction,* that is, they *may* conceivably be cognitively correct. Elsewhere, Rokeach and Eglash (20) have addressed themselves more specifically to this issue. The position taken there is that there are compelling theoretical reasons for drawing a distinction between strong beliefs rooted in dogmatic conviction and strong beliefs rooted in intellectual conviction. On the basis of such as a distinction, it may be added, a measure of intellectual conviction was constructed in which a number of statements give a clearly "wrong" reason for holding to the "right" belief. Lack of intellectual conviction, thus determined, was found to be directly related to measures of dogmatic conviction.

Dogmatism distinguished from rigidity. Another theoretical distinction which should be kept in mind is that between dogmatism and rigidity (19). The former construct refers to closed *systems* of sets or beliefs; the latter to single sets or beliefs which resist change. The difference between dogmatic and rigid thinking can perhaps be more clearly indicated by pointing to a recent experiment by Rokeach, McGovney, and Denny (23). Subjects are presented with a problem in which *(a)* three separate sets must first be overcome and *(b)* then integrated in order to reach a solution. Subjects scoring high on a rigidity scale are significantly slower than those scoring low in overcoming the three sets. On the other hand, subjects scoring high on the Dogmatism scale are significantly slower than those scoring low in time taken to integrate the new sets after the older sets had been overcome. These findings indicate not only that the distinction drawn between dogmatism and rigidity is a meaningful one but also that personality scales designed to measure dogmatism and rigidity are tapping different dimensions.

Dogmatism and anxiety. A comment on the motivational function served by dogmatism: In Section II we have made certain assumptions, some of which were reflected in the Dogmatism scale, regarding the motivational conditions which lead to dogmatism. In a factorial study by Rokeach and Fruchter (21) evidence is presented to the effect that tests designed to measure dogmatism and anxiety, along with other tests designed to measure self-rejection and paranoid tendencies, emerged together as part of a general anxiety factor. It may be stated further, for the sake of completeness, that in all samples tested the relation between dogmatism and anxiety is always positive.[23]

---

[23] For the seven samples of the present study the product-moment correlations between dogmatism and anxiety are, respectively, .44, .58,

Some methodological considerations. There are certain methodological problems to which attention should be drawn. It will be noted that the procedures employed in constructing the Dogmatism and Opinionation scales are very similar to those employed in the California research. For one thing all the items in the D and O scales, like those in the California scales, are worded in the same direction—agreement indicating highness on the variables in question. Thus, it is likely that response set (3) was operating to at least some extent in determining individual scores. It is worth pointing out, however, that in our preliminary research we used dogmatism items worded in the opposite direction—that is, disagreement indicating highness. Such items had to be discarded because they consistently failed to discriminate low from high dogmatic groups. Apparently, the subjects could "see through" such items relatively easily.

It should be pointed out too, in view of the fact that the Dogmatism and Opinionation scales have been pitted against the F and E scales in the present research, that it was actually advantageous to employ items all of which were worded negatively. The fact that systematic differences have been found between the two sets of scales despite their being equated for response set serves to strengthen rather than weaken the conclusions reached regarding such differences.

Another methodological problem has to do with the Opinionation scale. A major principle of test construction is that double-barreled items should be avoided. Yet the concept of Opinionation, by its very nature, necessarily involves the presentation of double-barreled statements. That these double-barreled statements have not hampered the goals of the present research is indicated by the reliabilities found for the total Opinionated scale and the even higher reliabilities found for the Left and Right subscales, and furthermore, by the manner in which these variables are differentially related to the other variables under consideration.

XI. SUMMARY AND CONCLUSIONS

The present research rests, first, on the assumption that the conceptualizations and measuring devices reported by Adorno *et al.* (1) have to do primarily with right, rather than with general, authoritarianism and intolerance. Second, the present research rests on an alternative set of theoretical formulations regarding the nature of political and religious dogmatism. These formulations are designed to deal with and measure *general* authoritarianism and *general* intolerance, regardless of specific ideological content.

We have presented herein a theoretical analysis of the cognitive structure, formal content, and function served by dogmatic thinking and believing. Guided by this analysis two scales were constructed: the Dogmatism and Opinionation scales. The former scale represents, by definition, a measure of (*a*) openness-closedness of

---

.64, .57, .52, .47, and .36. These findings suggest that those religious and political groups which are the most dogmatic should also turn out to to be the most anxious. To find out whether this is the case, anxiety means were calculated for the several religious groups in the Michigan State University and New York Colleges samples; and for the several political groups in the English Colleges I sample.

For the Michigan State University Catholics, Protestants and "Nones," the mean anxiety scores are 83.07, 83.55, and 71.87, respectively ($SD = 19.15$, 25.17, and 18.91, respectively). The differences between Catholics and "Nones," and between Protestants and "Nones" are significant at the .05 level.

For the New York Catholics, Protestants, Jews, and "Nones," the mean anxiety scores are 85.02, 71.54, 79.89, and 69.50, respectively ($SD = 28.64$, 24.73, 25.44, and 9.78 respectively). Significant at the .05 level are the differences between Catholics and Protestants, Catholics and "Nones," and between Jews and "Nones."

For the English Conservatives, Liberals, Attleeites, Bevanites, and Communists the mean anxiety scores are 81.02, 81.27, 80.19, 90.37, and 73.46, respectively. ($SD = 25.68$, 20.35, 24.22, 21.43, and 21.97 respectively.) The difference between Bevanites and Communists is significant at the .05 level.

It is clear that the most dogmatic groups are not necessarily the most anxious. As can be seen, the "None" groups and the Communists score lower on anxiety than do any of the other groups.

Further research along these lines should prove rewarding.

cognitive systems, (*b*) general authoritarianism, and (*c*) general intolerance. The Opinionation scale was designed to measure only *c* above, namely, the extent to which there is intolerance or qualified tolerance toward others according to whether others disagree or agree with our own beliefs. Furthermore, the Opinionation scale is so constructed so that it is possible to obtain measures of left and right opinionation as well as total opinionation.

The subjects were students attending universities in Michigan, New York, and England. Also tested was a group of automobile workers in England. The corrected reliabilities of successive editions of the Dogmatism scale ranged from .70 to .91; the corrected reliabilities of successive editions of the Opinionation scale ranged from .67 to .75.

To determine the validity of our conceptual representation of dogmatism and the operational measurement thereof by means of the Dogmatism and Opinionation scales three major hypotheses were tested. These three hypotheses were designed (*a*) to test different facets of our conceptual representation of the dogmatism, and (*b*) to pit systematically the present formulations and scales against those presented by Adorno *et al.* (1), in order to determine the differential properties of the two sets.

Hypothesis A

This is to the effect that the Dogmatism scale represents a measure of *general* authoritarianism while, in contrast, the F scale measures *right* authoritarianism.

1. We find first that the Dogmatism (D) scale correlates relatively highly with the F scale ($r = .54$ to .77). These correlations are about as high as those generally reported between F and E. The conclusion which may be drawn thus far is that both D and F represent equally good measures of authoritarianism.

2. But the Dogmatism scale seems to differ from the F scale in certain respects. The correlations between D and E are consistently smaller than the correlations between F and E. Even when ethnocentrism is held constant, by the method of partial correlation, the correlations between D and F are reduced but slightly. From these findings it may be concluded that the Dogmatism scale measures individual differences in authoritarianism independent of highness or lowness in ethnic attitudes.

3. The Dogmatism scale correlates negligibly with liberalism-conservatism, and in every group sampled these correlations are considerably smaller than those between the F scale and liberal-conservatism. These findings, along with those mentioned above, suggest that dogmatism comes closer to measuring general authoritarianism than does the F scale. They indicate too that the phenomenon of dogmatism has been found with approximately equal frequency along all positions of the left-to-right continuum.

4. Further analysis of each of the individual items of the F scale reveals that approximately one-third of the F items measure right authoritarianism, another third general authoritarianism, and the remaining third something in between right and general authoritarianism. The net effect is a scale biased in the direction of measuring primarily right authoritarianism.

Hypothesis B

This is to the effect that dogmatism, as conceived and measured, is related both to left-of-center and right-of-center forms of intolerance (as measured by the Opinionation scale) while, in contrast, the F and E scales are conceived and constructed in such a way that they tap primarily right-of-center forms of intolerance.

1. In all groups dogmatism correlates positively with both left and right opinionation despite the fact that left and right opinionation are negligibly or, more often, negatively related to each other. In contrast, both the F and E scales correlate relatively highly with right opinionation but negligibly or negatively with left opinionation.

2. In every group the Dogmatism scale correlates more highly with total opinionation than with left or right opinionation. This is not so with the F, E, and P.E.C. scales. In almost all instances these latter three variables correlate *less* with total opinionation than with right opinionation, due to the generally negative correlations between each of these three variables and left opinionation.

Hypothesis C

This hypothesis is to the effect that the existence of strong group pressures to commitment (e.g. the Communist Party, the Catholic Church) will lead to greater dogmatism and opinionation, regardless of whether the commitment is oriented to left- or right-of-center ideologies. In contrast, it is hypothesized that high F- and E-scale scores will be manifest only in the case of right-oriented commitments.

1. Catholic and Communist groups are found to diverge sharply from each other with respect to their performance on the California F and E scales. The Catholic group is highest of all religious groups on both F and E; the Communist group is lowest of all political groups on both F and E.

2. Even though the Catholic and Communist groups score at opposite poles on F and E, these groups generally score highest on dogmatism and opinionation.

These findings are in accord with Hypothesis C and indicate that while the F and E scales measure right authoritarianism and intolerance, the Dogmatism and Opinionation scales measure general authoritarianism and intolerance.

REFERENCES

1. ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality.* New York: Harper, 1950.
2. BERGER, E. M. The relation between expressed acceptance of self and expressed acceptance of others. *J. abnorm. soc. Psychol.,* 1952, **47,** 778-782.
3. COHN, T. S. Is the F scale indirect? *J. abnorm. soc. Psychol.,* 1952, **47,** 732.
4. CHRISTIE, R., & JAHODA, MARIE. *Studies in the scope and method of "the authoritarian personality."* Glencoe, Illinois: Free Press, 1954.
5. CHRISTIE, R. Eysenck's treatment of the personality of communists. *Psychol. Bull.,* in press.
6. EDWARDS, A. L. *Statistical methods for the behavioral sciences.* New York: Rinehart, 1954.
7. EYSENCK, H. J. *The psychology of politics.* London: Routledge & Kegan Paul, 1954.
8. FRANK, L. K. Time perspectives. *J. soc. Phil.,* 1939, **4,** 293-312.
9. FROMM, E. *Man for himself.* New York: Farrar & Rinehart, 1947.
10. GOODSTEIN, L. D. Intellectual rigidity and social attitudes *J. abnorm. soc. Psychol.,* 1953, **48,** 345-353.
11. GOUGH, H. G., & SANFORD, R. N. *Rigidity as a psychological variable.* Berkeley: Univer. of Calif. Inst. of Pers. Assessment and Res., 1952 (mimeographed).
12. HATHAWAY, S. R., & MCKINLEY, J. C. The Minnesota Multiphasic Personality Inventory. New York: Psychological Corp., 1943.
13. HOFFER, E. *The true believer.* New York: Harper, 1951.
14. INKELES, A., & LEVINSON, D. J. National character: The study of modal personality and sociocultural systems. In G. Lindzey, (Ed.), *Handbook of social psychology.* Cambridge: Addison-Wesley, 1954.
15. LEWIN, K. *Field theory in social science.* New York: Harper, 1951.
16. MASLOW, A. H. *Motivation and personality.* New York: Harper, 1954.
17. ROKEACH, M. A method for studying individual differences in "narrow-mindedness." *J. Pers.,* 1951, **20,** 219-233.
18. ROKEACH, M. Dogmatism and opinionation on the left and on the right. *Amer. Psychologist,* 1952, **7,** 310. (Abstract)
19. ROKEACH, M. The nature and meaning of dogmatism. *Psychol. Rev.,* 1954, **61,** 194-204.
20. ROKEACH, M., & EGLASH, A. A scale for measuring intellectual conviction. *J. soc. Psychol,* in press.
21. ROKEACH, M., & FRUCHTER, B. A. factorial study of dogmatism and related concepts. *J. abnorm. soc. Psychol.,* in press.
22. ROKEACH, M., & HANLEY, C. Eysenck's tender-mindedness dimension: A critique. *Psychol. Bull.* 1956, **53,** 169-176.
23. ROKEACH, M., MCGOVNEY, W. C., & DENNY, M. R. A distinction between dogmatic and rigid thinking. *J. abnorm. soc. Psychol.,* 1955, **51,** 87-93.
24. SANFORD, R. N. Recent developments in connection with the investigation of the authoritarian personality. *Sociol. Rev.,* 1954, **2,** 11-33.
25. SULLIVAN, P. L., & ADELSON, J. Ethnocentrism and misanthropy. *J. abnorm. soc. Psychol.,* 1953, **49,** 246-250.
26. TAYLOR, I. A. Perceptual closure and extreme social attitudes. *Amer. Psychologist,* 1954, **9,** 482. (Abstract)
27. WELCH, G. S. An anxiety index and an internalization ratio for the MMPI. *J. consult. Psychol.,* 1952, **16,** 65-72.

(Accepted for publication December 5, 1955)

Vol. 70, No. 19 Whole No. 426, 1956
Psychological Monographs: General and Applied

BUREAU OF EDNL. & LIBRAR

Attitudes of Young Children Toward Their Peers as Related to Certain Characteristics of Their Siblings[1, 2]

HELEN L. KOCH
University of Chicago

I. PROBLEM, SUBJECTS, AND PROCEDURE

THIS study is a part of a more extensive one concerned with the relation between some of the social attitudes of the five- and six-year-old in two-child families and the variables of sex, ordinal position, and age-difference of the child's sibling. Sex differences were also investigated. The relations between the sex and sibling variables just mentioned and the following dependent variables were explored: friendliness to children, gregariousness, leadership, popularity, kindness, quarrelsomeness, revengefulness, cruelty, tendency to tease, uncooperativeness with peers, jealousy, competitiveness, sportmanship, selfishness, insistence on rights, and exhibitionism.

The subjects of the study were 384 white, "normal," five- and six-year-olds from native-born, intact, urban, two-child families. No adopted or stepchildren were included in the population. There were 48 children in each of the following groups: males with an older brother, males with a younger brother, males with an older sister, males with a younger sister, females with an older brother, females with a younger brother, females with an older sister and females with a younger sister. Each of these groups of 48 was divided further into three subgroups of 16 in which the age difference between the sibs was, respectively, less than two years, two to four years, and four to six years. The arrangement includes, then, three spacings, two ordinal positions, child subjects of two sexes and their siblings of two sexes (see Table 1). When the sib is younger, the age-difference variable includes not only a maturity difference but a length-of-contact variable.

The children were mostly public school children, only one private school system contributing subjects. Each of the subgroups of 16 was matched with every other, individual by individual, with respect to age of child (Table 2) as well as socioeconomic level of the

---

[1] The investigation was supported, in part, by a grant from the National Institute of Mental Health, of the National Institutes of Health, Public Health Service, and in part, by the Sonia Shankman Orthogenic School Research Fund, University of Chicago.

[2] The author wishes to acknowledge the help received from, and to express her gratitude to, the teachers and principals of the following schools in Chicago which participated in the study: Bennett School, Bradwell School, Field School, Holmes School, Gunsalus School, Luella School, Perry School, Schmid School, Sullivan School, Wentworth School, University of Chicage Laboratory Schools and the University of Chicago Nursery School. The author is also much indebted to Dr. Grace Munson and Dr. Thelma Thurstone, both formerly Directors of the Bureau of Child Study, Chicago Public Schools. Dr. Helen Heath did most of the statistical work for the study. We acknowledge with gratitude her discriminating and loyal service.

Bureau Ednl. Psy. Research
DAVID H ING COLLEGE
Dated ....
Accs No ....

TABLE 1

Number of Subjects in Various Groups

Description of Group			Age Difference Between Siblings in Months			Total
Subject		Sibling				
Sex	Ordinal Position	Sex	7–24	25–48	49–72	
M	2	M	16	16	16	48
M	2	F	16	16	16	48
F	2	M	16	16	16	48
F	2	F	16	16	16	48
M	1	M	16	16	16	48
M	1	F	16	16	16	48
F	1	M	16	16	16	48
F	1	F	16	16	16	48
M	2	M or F	32	32	32	96
F	2	M or F	32	32	32	96
M	1	M or F	32	32	32	96
F	1	M or F	32	32	32	96
M or F	2	M	32	32	32	96
M or F	2	F	32	32	32	96
M or F	1	M	32	32	32	96
M or F	1	F	32	32	32	96
M	1 or 2	M	32	32	32	96
M	1 or 2	F	32	32	32	96
F	1 or 2	M	32	32	32	96
F	1 or 2	F	32	32	32	96
M	1 or 2	M or F	64	64	64	192
F	1 or 2	M or F	64	64	64	192
M or F	1 or 2	M	64	64	64	192
M or F	1 or 2	F	64	64	64	192
M or F	1	M or F	64	64	64	192
M or F	2	M or F	64	64	64	192
M or F	1 or 2	M or F	128	128	128	384

occupation of the subject's father (Table 3) and of the neighborhood of residence of the family (Table 4). Occupation classification was based on the Minnesota Scale (8); and the neighborhood classification, on the Chicago Community Inventory (13).

Matching could not, of course, be exact. Ninety-eight per cent of the age matchings were within six months; 93 per cent of the occupation matchings, within one level; and 88 per cent of the neighborhood, within one scale rank. Level 4 on the occupation scale was omitted because this is the farming group, a class not represented in the population we studied. Tables 2 to 4 show the composition of our various basic subgroups of 16; and Table 1, the total number of cases in each of the various subgroups or subgroup combinations.

The measures of the traits studied were teachers' ratings. About half of the rating scales used were Fels Research Institute Child Behavior Scales (14); while the others were taken from the California Behavior Inventory for Nursery School Children (4) but were handled as graphic scales, as the Fels Scales are normally. To designate the traits measured with the help of the Fels Scale we have placed a star after the captions in the table of means (Tables 5 to 13). The unstarred scales, then, are from the California Inventory. Since the teachers saw the children only at school and with their classmates, the ratings describe chiefly the attitudes our subjects showed toward their peers at school.

The method of indicating an assessment on the line scales was merely a check mark at the appropriate point, the descriptions of various degrees of the trait serving merely as landmarks and not being evenly spaced. The checkings on the line scale were then converted with the use of a grid into ratings on a nine-point scale. These ratings, in turn, were normalized on the basis of a population of about 500 five- and six-year-olds of which our 384 subjects were a major part. Four was taken as the mean score. The ratings for the 24 sub-

TABLE 2

MEANS AND SIGMAS OF THE AGES OF SUBJECTS AND SIBLINGS IN THE TWENTY-FOUR BASIC GROUPS

Description of Group			Age in Months					
Subject		Sibling	Subject		Sibling		Sibling Age Difference	
Sex	Ordinal Position	Sex	Mean	Sigma	Mean	Sigma	Mean	Sigma
I. Sibling Age Difference 7–24 Months								
M	2	M	71.4	6.1	89.6	6.8	18.2	3.6
M	2	F	69.3	6.2	87.8	7.6	18.4	3.3
F	2	M	70.4	7.7	89.3	7.5	18.9	4.1
F	2	F	71.3	5.2	90.6	6.3	19.3	4.1
M	1	M	68.4	6.1	50.1	6.0	18.3	3.4
M	1	F	71.4	7.4	53.3	9.1	18.1	4.2
F	1	M	70.4	6.3	52.7	7.1	17.8	4.2
F	1	F	70.9	5.7	51.4	6.9	19.5	3.6
II. Sibling Age Difference 25–48 Months								
M	2	M	70.9	6.1	105.6	6.8	34.8	6.7
M	2	F	71.8	6.3	108.9	8.8	37.1	6.4
F	2	M	72.2	6.5	107.7	10.0	35.4	5.8
F	2	F	72.8	5.4	107.9	8.4	35.1	6.3
M	1	M	71.1	6.6	35.0	8.3	36.1	6.6
M	1	F	70.9	6.0	32.3	8.8	38.6	6.9
F	1	M	71.8	7.2	37.9	11.6	33.9	7.1
F	1	F	72.4	5.6	37.6	6.3	34.9	6.0
III. Sibling Age Difference 49–72 Months								
M	2	M	68.9	6.2	129.6	8.0	60.7	7.8
M	2	F	73.1	7.0	131.8	10.8	58.6	6.0
F	2	M	71.9	4.3	125.7	7.5	53.8	5.0
F	2	F	67.6	5.1	123.7	6.6	56.1	5.1
M	1	M	72.8	6.7	18.2	6.9	54.5	5.8
M	1	F	72.3	7.5	16.8	6.6	55.6	5.4
F	1	M	70.6	5.5	14.8	7.2	55.8	4.5
F	1	F	73.4	5.7	17.1	6.0	56.2	5.8

TABLE 3

DISTRIBUTION OF SUBJECTS IN TWENTY-FOUR BASIC GROUPS IN TERMS OF FATHER'S OCCUPATION

Description of Group			Age Difference Between Siblings 7 to 24 Months						Age Difference Between Siblings 25 to 48 Months						Age Difference Between Siblings 49 to 72 Months					
Subject		Sibling	Class of Father's Occupation																	
Sex	Ordinal Position	Sex	I	II	III	V	VI	VII	I	II	III	V	VI	VII	I	II	III	V	VI	VII
			Number of Subjects																	
M	2	M	5	—	4	7	—	—	5	3	4	3	—	1	—	5	4	6	—	1
M	2	F	2	3	5	5	1	—	5	2	5	4	—	—	2	5	3	4	2	—
F	2	M	3	3	3	7	—	—	5	5	2	2	2	—	3	5	4	4	—	—
F	2	F	4	4	4	3	1	—	4	4	3	4	1	—	2	3	4	4	3	—
M	1	M	4	1	5	4	2	—	6	2	2	6	—	—	5	4	4	2	1	—
M	1	F	5	1	2	7	1	—	6	2	2	5	1	—	6	2	3	1	3	1
F	1	M	3	2	5	6	—	—	5	5	1	3	2	—	2	4	4	4	1	1
F	1	F	3	4	4	5	—	—	3	4	6	3	—	—	4	4	3	4	—	1

groups were checked by the Bartlett test (6, pp. 195-200) to determine whether it was reasonable to assume the groups were drawn from populations with a common variance.

The teachers, most of them very experienced, were coached in a conference or two on how to make the ratings. Since children from 12 schools were used, as many as 12 teachers in a school sometimes made the ratings, and those assessments by a given teacher were widely spread through the groups, it seems unlikely that differences in the standards of individual judges are responsible for any of our group differences. Since the

TABLE 4

DISTRIBUTION OF SUBJECTS IN THE TWENTY-FOUR BASIC GROUPS IN TERMS OF NEIGHBORHOOD OF RESIDENCE

Description of Group			Age Difference Between Siblings 7 to 24 Months					Age Difference Between Siblings 25 to 48 Months					Age Difference Between Siblings 49 to 72 Months				
Subject		Sibling	Socioeconomic Classification of Neighborhood of Residence														
Sex	Ordinal Position	Sex	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
			Number of Subjects														
M	2	M	8	3	3	1	1	6	5	2	3	—	11	3	1	—	1
M	2	F	8	3	—	3	2	10	1	4	—	1	9	2	1	4	—
F	2	M	8	5	2	1	—	11	—	2	2	1	11	3	1	1	—
F	2	F	11	3	—	2	—	8	2	2	4	—	7	4	2	2	1
M	1	M	9	1	1	3	2	9	1	5	1	—	10	1	3	1	1
M	1	F	9	4	3	—	—	8	1	5	2	—	8	1	5	2	—
F	1	M	8	1	3	2	2	10	4	2	—	—	12	1	1	2	—
F	1	F	6	6	1	2	1	8	2	5	1	—	14	—	—	1	1

schools did not contribute to the groups in the same proportion, a school standard variable may have colored our results to some extent. The matching of the groups in socioeconomic status, doubtless, kept the effect of such a variable to a minimum. Furthermore, most of the teachers had taught in a variety of schools; so their ideas of child behavior were not built up entirely on the basis of the child population in the schools at which they were teaching at the time they cooperated in our study. We have also counted on the relative specificness of the descriptions of the levels in the various scales as protection (partial if not complete) against possible differences in school standards.

There are still other selective influences, of course, that may be buried in the data. Age of mother is one of these; family attitude, another. It is probable

TABLE 5

MEANS OF THE NORMALIZED RATINGS FOR THE VARIOUS SUBGROUPS: FRIENDLINESS TO PEERS* AND GREGARIOUSNESS*

Description of Group			Friendliness to Peers*				Gregariousness*			
Subject		Sibling	Age Difference Between Siblings in Months							
Sex	Ordinal Position	Sex	7–24	25–48	49–72	Total	7–24	25–48	49–72	Total
			Score Means				Score Means			
M	2	M	3.33	3.68	4.29	3.76	3.66	4.10	4.47	4.08
M	2	F	3.58	3.62	3.74	3.64	3.67	3.62	3.70	3.66
F	2	M	4.12	4.25	3.98	4.12	4.06	4.16	3.76	3.99
F	2	F	4.04	3.82	4.20	4.02	3.96	3.82	3.99	3.92
M	1	M	3.82	3.80	4.30	3.97	3.58	3.93	4.06	3.86
M	1	F	2.98	4.38	4.03	3.80	3.33	4.36	4.18	3.96
F	1	M	4.22	4.07	4.32	4.20	4.37	4.11	4.39	4.29
F	1	F	4.10	4.08	4.07	4.08	4.07	4.02	4.31	4.13
M	2	M or F	3.45	3.65	4.02	3.70	3.67	3.86	4.08	3.87
F	2	M or F	4.08	4.03	4.09	4.07	4.01	3.99	3.88	3.96
M	1	M or F	3.40	4.09	4.16	3.88	3.45	4.15	4.12	3.91
F	1	M or F	4.16	4.08	4.20	4.14	4.22	4.06	4.35	4.21
M or F	2	M	3.72	3.97	4.14	3.94	3.86	4.13	4.12	4.04
M or F	2	F	3.81	3.72	3.97	3.83	3.81	3.72	3.85	3.79
M or F	1	M	4.02	3.94	4.31	4.09	3.97	4.02	4.23	4.08
M or F	1	F	3.54	4.23	4.05	3.[illegible]4	3.70	4.19	4.25	4.04
M	1 or 2	M	3.57	3.74	4.30	3.87	3.62	4.02	4.27	3.97
M	1 or 2	F	3.28	4.00	3.88	3.72	3.50	4.00	3.94	3.81
F	1 or 2	M	4.17	4.16	4.15	4.16	4.22	4.13	4.08	4.14
F	1 or 2	F	4.07	3.95	4.14	4.05	4.01	3.92	4.15	4.03
M	1 or 2	M or F	3.43	3.87	4.09	3.80	3.56	4.01	4.10	3.89
F	1 or 2	M or F	4.12	4.05	4.14	4.11	4.11	4.02	4.12	4.08
M or F	1 or 2	M	3.87	3.95	4.22	4.02	3.92	4.08	4.17	4.06
M or F	1 or 2	F	3.68	3.97	4.01	3.89	3.76	3.96	4.05	3.92
M or F	2	M or F	3.77	3.84	4.05	3.89	3.84	3.93	3.98	3.92
M or F	1	M or F	3.78	4.08	4.18	4.01	3.84	4.11	4.24	4.06
M or F	1 or 2	M or F	3.77	3.96	4.12	3.95	3.84	4.02	4.11	3.99

* From reference 14.

TABLE 6

Means of the Normalized Ratings for the Various Subgroups: Popularity and Leadership*

Description of Group			Popularity				Leadership*			
Subject		Sibling	Age Difference Between Siblings in Months							
Sex	Ordinal Position	Sex	7–24	25–48	49–72	Total	7–24	25–48	49–72	Total
			Score Means				Score Means			
M	2	M	3.70	4.01	4.29	4.00	3.80	3.84	4.45	4.03
M	2	F	3.74	3.81	3.74	3.76	3.71	3.73	3.69	3.71
F	2	M	4.22	4.14	4.02	4.12	4.06	4.24	4.19	4.16
F	2	F	3.79	3.35	4.31	3.82	3.86	3.77	4.25	3.96
M	1	M	3.72	3.64	4.14	3.84	3.69	3.52	4.39	3.87
M	1	F	3.56	4.20	3.88	3.88	3.57	4.28	4.33	4.06
F	1	M	4.20	4.26	4.18	4.21	4.45	4.36	4.30	4.37
F	1	F	4.04	4.30	4.28	4.21	3.69	4.06	3.95	3.90
M	2	M or F	3.72	3.91	4.01	3.88	3.76	3.79	4.07	3.87
F	2	M or F	4.00	3.74	4.16	3.97	3.96	4.01	4.22	4.06
M	1	M or F	3.64	3.92	4.01	3.86	3.63	3.90	4.36	3.96
F	1	M or F	4.12	4.28	4.23	4.21	4.07	4.21	4.13	1.14
M or F	2	M	3.96	4.08	4.15	4.06	3.93	4.04	4.32	4.10
M or F	2	F	3.77	3.58	4.02	3.79	3.79	3.75	3.97	3.84
M or F	1	M	3.96	3.95	4.16	4.02	4.07	3.94	4.35	4.12
M or F	1	F	3.80	4.25	4.08	4.04	3.63	4.17	4.14	3.98
M	1 or 2	M	3.71	3.83	4.22	3.92	3.75	3.68	4.42	3.95
M	1 or 2	F	3.65	4.00	3.81	3.82	3.64	4.01	4.01	3.89
F	1 or 2	M	4.21	4.20	4.10	4.17	4.25	4.30	4.25	4.27
F	1 or 2	F	3.92	3.83	4.29	4.01	3.78	3.92	4.10	3.93
M	1 or 2	M or F	3.68	3.92	4.01	3.87	3.69	3.84	4.22	3.92
F	1 or 2	M or F	4.06	4.02	4.20	4.09	4.02	4.11	4.17	4.10
M or F	1 or 2	M	3.96	4.02	4.16	4.04	4.00	3.99	4.34	4.11
M or F	1 or 2	F	3.78	3.92	4.05	3.92	3.71	3.96	4.06	3.91
M or F	2	M or F	3.86	3.83	4.09	3.93	3.86	3.90	4.15	3.97
M or F	1	M or F	3.88	4.10	4.12	4.03	3.85	4.06	4.24	4.05
M or F	1 or 2	M or F	3.87	3.97	4.10	3.98	3.85	3.98	4.20	4.01

* Starred scale from reference 14, unstarred from reference 4.

that the mothers of our second-borns were slightly older than the mothers of our first-borns, for instance. We had no way of checking on this. (See 10 and 11 for further comment on selective influences.)

We are presenting variance data for the total population only (see Tables 14-22). Significant specific relations in the spacing subgroups which are not spelled out by the variance data offered will be described as we go along. The *F*'s are given in the tables only if statistically significant, or if approaching statistical significance. With 1 and 345 degrees of freedom, an *F* of 3.86 is needed for significance at the 5 per cent point and of 6.70 for significance at the 1 per cent point; while with 2 and 345 degrees of freedom, the *F*'s necessary for comparable *p*'s are, respectively, 3.02 and 4.06. In computing the *F*'s the residual variance was used as the error term (see Tables 14-22). In our data no single variable effect, since the degrees of freedom are small, can be shown to operate over and above interaction effects when these are significant.

When necessary to judge significance, *t*'s have been computed; but the *t*'s are not presented here in the interests of economy. Relations described can be assumed, unless we specify to the contrary, to be significant at the 5 per cent point or better, a two-tailed test having been employed. Minimal differences between the smaller group means of .45 to .50 tend to be necessary for significance, though, of course, no specific quantity can be named (Tables 5-13). Should the reader wish some idea of the likely reliabilities of the measures used, we can refer him to the monograph by Richards and Powell (14) for a rough notion of

TABLE 7

MEANS OF THE NORMALIZED RATINGS FOR THE VARIOUS SUBGROUPS: KINDNESS*

Description of Group			Kindness*			
Subject		Sibling	Age Difference Between Siblings in Months			
Sex	Ordinal Position	Sex	7–24	25–48	49–72	Total
			Score Means			
M	2	M	3.78	3.75	4.12	3.88
M	2	F	3.98	4.18	3.46	3.87
F	2	M	4.14	4.13	4.02	4.09
F	2	F	4.20	3.67	4.07	3.98
M	1	M	3.77	3.28	4.34	3.80
M	1	F	3.92	3.85	4.01	3.93
F	1	M	4.30	3.86	4.03	4.06
F	1	F	3.81	4.03	3.83	3.89
M	2	M or F	3.88	3.96	3.79	3.88
F	2	M or F	4.17	3.90	4.04	4.04
M	1	M or F	3.84	3.56	4.18	3.86
F	1	M or F	4.05	3.95	3.93	3.98
M or F	2	M	3.96	3.94	4.07	3.99
M or F	2	F	4.09	3.92	3.76	3.92
M or F	1	M	4.04	3.57	4.18	3.93
M or F	1	F	3.86	3.94	3.92	3.91
M	1 or 2	M	3.78	3.51	4.23	3.84
M	1 or 2	F	3.95	4.02	3.74	3.90
F	1 or 2	M	4.22	3.99	4.02	4.08
F	1 or 2	F	4.00	3.85	3.95	3.93
M	1 or 2	M or F	3.86	3.76	3.98	3.87
F	1 or 2	M or F	4.11	3.92	3.98	4.01
M or F	1 or 2	M	4.00	3.75	4.13	3.96
M or F	1 or 2	F	3.98	3.93	3.84	3.92
M or F	2	M or F	4.02	3.93	3.92	3.96
M or F	1	M or F	3.95	3.76	4.05	3.92
M or F	1 or 2	M or F	3.99	3.84	3.98	3.94

* From reference 14.

TABLE 8

MEANS OF THE NORMALIZED RATINGS FOR THE VARIOUS SUBGROUPS: QUARRELSOMENESS* AND REVENGEFULNESS

Description of Group			Quarrelsomeness*				Revengefulness			
Subject		Sibling	Age Difference Between Siblings in Months							
Sex	Ordinal Position	Sex	7–24	25–48	49–72	Total	7–24	25–48	49–72	Total
			Score Means				Score Means			
M	2	M	4.56	4.14	4.47	4.39	4.48	3.94	4.12	4.18
M	2	F	4.07	4.10	4.31	4.16	4.27	4.03	4.21	4.17
F	2	M	3.63	4.03	3.53	3.73	3.68	4.10	3.95	3.91
F	2	F	3.88	3.99	3.56	3.81	3.79	4.08	3.83	3.90
M	1	M	3.68	4.49	3.89	4.02	4.11	4.51	3.97	4.20
M	1	F	4.04	4.64	4.34	4.34	4.16	4.59	4.00	4.25
F	1	M	4.40	4.03	4.31	4.25	3.72	3.76	4.00	3.83
F	1	F	3.89	3.91	3.80	3.87	3.95	3.98	4.00	3.98
M	2	M or F	4.32	4.12	4.39	4.28	4.37	3.99	4.16	4.17
F	2	M or F	3.76	4.01	3.54	3.77	3.73	4.09	3.89	3.91
M	1	M or F	3.86	4.56	4.12	4.18	4.14	4.55	3.99	4.22
F	1	M or F	4.14	3.97	4.06	4.06	3.83	3.87	3.99	3.90
M or F	2	M	4.10	4.08	4.00	4.06	4.08	4.02	4.04	4.04
M or F	2	F	3.98	4.04	3.93	3.98	4.03	4.06	4.02	4.04
M or F	1	M	4.04	4.26	4.10	4.13	3.91	4.14	3.98	4.01
M or F	1	F	3.97	4.28	4.07	4.10	4.06	4.28	4.00	4.11
M	1 or 2	M	4.12	4.31	4.18	4.20	4.29	4.23	4.04	4.19
M	1 or 2	F	4.06	4.37	4.32	4.25	4.21	4.31	4.11	4.21
F	1 or 2	M	4.01	4.03	3.92	3.99	3.70	3.93	3.98	3.87
F	1 or 2	F	3.89	3.95	3.68	3.84	3.87	4.03	3.92	3.94
M	1 or 2	M or F	4.09	4.34	4.25	4.23	4.25	4.27	4.08	4.20
F	1 or 2	M or F	3.95	3.99	3.80	3.91	3.78	3.98	3.95	3.90
M or F	1 or 2	M	4.07	4.17	4.05	4.10	3.99	4.08	4.01	4.03
M or F	1 or 2	F	3.97	4.16	4.00	4.04	4.04	4.17	4.01	4.08
M or F	2	M or F	4.04	4.06	3.96	4.02	4.05	4.04	4.03	4.04
M or F	1	M or F	4.00	4.27	4.09	4.12	3.98	4.21	3.99	4.06
M or F	1 or 2	M or F	4.02	4.17	4.03	4.07	4.02	4.12	4.01	4.05

* Starred scale from reference 14, unstarred from reference 4.

those that may be expected in the case of the Fels items, and to (5) for those likely in the case of the California Inventory items. We have no reliability data to offer which are based on our material.

We shall often refer to the groups in code—e.g., M2F [25-48]. The first letter refers to the sex of the children in the group and the first number to their ordinal position, while the second letter refers to the sex of their sibs. The numbers in brackets indicate the range, in months, of the age difference between the children and their sibs. Thus M2F [25-48] designates the group of males each of whom has a sister within two to four years older.

II. RESULTS

Since the detail may seem overwhelming, the results will first be stated here in outline form. In the next section major relations will be singled out for discussion, and some group characterizations will be attempted. Let us also say that, because the interrelations among our variables are so complicated, we have under the various sections presented below sometimes given a refocused description of a relation mentioned under an earlier section.

Sib-Age-Disparity Group Differences

1. The variable, age difference between sibs, seems related in a simple way to few of our dependent variables, but appeared in many significant interactions with other independent variables, especially with sex and ordinal position.

2. Sib spacing and the traits we studied showed little correlation in the case of first-born girls or in the case of second-born boys with a sister.

3. The two-to-four-year spacing, as Lasko (12), Foster (7) and Stendler (20) surmised, seemed, according to our findings also, to be a particularly stimulating and stressful one, especially when the sibs were opposite in sex.

4. Although girls tended not to show the trend, boys, with the exception of the M2F's, were, as sib-age disparity widened, rated higher, on the average, in friendliness to peers, gregar-

TABLE 9

MEANS OF THE NORMALIZED RATINGS FOR THE VARIOUS SUBGROUPS: CRUELTY* AND TENDENCY TO FIND FAULT

Description of Group			Cruelty*				Tendency to Find Fault			
Subject		Sibling	Age Difference Between Siblings in Months							
Sex	Ordinal Position	Sex	7–24	25–48	49–72	Total	7–24	25–48	49–72	Total
			Score Means				Score Means			
M	2	M	4.37	3.83	4.25	4.15	4.01	3.95	4.29	4.08
M	2	F	4.32	4.19	4.48	4.33	3.82	4.02	3.86	3.90
F	2	M	3.61	3.95	4.14	3.90	3.85	3.60	4.26	3.90
F	2	F	3.70	4.08	3.78	3.85	3.92	4.08	4.19	4.06
M	1	M	3.98	4.57	4.02	4.19	3.68	4.12	4.01	3.94
M	1	F	4.06	4.35	4.04	4.15	4.16	4.34	4.20	4.23
F	1	M	4.25	4.15	3.85	4.08	4.19	3.97	4.52	4.22
F	1	F	4.19	4.03	4.00	4.07	3.59	4.24	3.88	3.90
M	2	M or F	4.35	4.01	4.37	4.24	3.92	3.99	4.08	3.99
F	2	M or F	3.66	4.02	3.96	3.88	3.88	3.84	4.22	3.98
M	1	M or F	4.02	4.46	4.03	4.17	3.92	4.23	4.10	4.08
F	1	M or F	4.22	4.09	3.93	4.08	3.89	4.10	4.20	4.06
M or F	2	M	3.99	3.89	4.20	4.03	3.93	3.77	4.27	3.99
M or F	2	F	4.01	4.13	4.13	4.09	3.87	4.05	4.03	3.98
M or F	1	M	4.12	4.36	3.94	4.14	3.93	4.04	4.26	4.08
M or F	1	F	4.13	4.19	4.02	4.11	3.87	4.29	4.04	4.07
M	1 or 2	M	4.18	4.20	4.14	4.17	3.84	4.04	4.15	4.01
M	1 or 2	F	4.19	4.27	4.26	4.24	3.99	4.18	4.03	4.07
F	1 or 2	M	3.93	4.05	4.00	3.99	4.02	3.78	4.39	4.06
F	1 or 2	F	3.95	4.05	3.89	3.96	3.75	4.16	4.04	3.98
M	1 or 2	M or F	4.18	4.23	4.20	4.21	3.92	4.11	4.09	4.04
F	1 or 2	M or F	3.94	4.05	3.94	3.98	3.88	3.97	4.21	4.02
M or F	1 or 2	M	4.06	4.12	4.07	4.08	3.93	3.91	4.27	4.04
M or F	1 or 2	F	4.07	4.16	4.08	4.10	3.87	4.17	4.03	4.02
M or F	2	M or F	4.00	4.01	4.16	4.06	3.90	3.91	4.15	3.99
M or F	1	M or F	4.12	4.27	3.98	4.12	3.90	4.17	4.15	4.07
M or F	1 or 2	M or F	4.06	4.14	4.07	4.09	3.90	4.04	4.15	4.03

* Starred scale from reference 14, unstarred from reference 4.

TABLE 10

Means of the Normalized Ratings for the Various Subgroups: Tendency to Tease and Exhibitionism

Description of Group			Tendency to Tease				Exhibitionism			
Subject		Sibling	Age Difference Between Siblings in Months							
Sex	Ordinal Position	Sex	7–24	25–48	49–72	Total	7–24	25–48	49–72	Total
			Score Means				Score Means			
M	2	M	3.83	4.01	4.27	4.04	3.87	4.38	4.17	4.14
M	2	F	4.01	4.35	4.96	4.44	4.27	3.86	4.30	4.14
F	2	M	3.41	3.98	3.90	3.76	3.83	3.84	3.92	3.86
F	2	F	3.58	4.04	3.86	3.83	3.77	3.96	3.64	3.79
M	1	M	3.98	4.41	4.00	4.13	4.17	4.15	3.90	4.07
M	1	F	3.88	4.77	4.33	4.33	3.78	4.76	4.34	4.29
F	1	M	3.95	3.83	4.18	3.98	4.04	4.10	4.53	4.22
F	1	F	3.84	4.13	3.88	3.95	3.92	3.68	3.60	3.73
M	2	M or F	3.92	4.18	4.61	4.24	4.07	4.12	4.24	4.14
F	2	M or F	3.49	4.01	3.88	3.79	3.80	3.90	3.78	3.83
M	1	M or F	3.93	4.59	4.16	4.23	3.97	4.46	4.12	4.18
F	1	M or F	3.89	3.98	4.03	3.97	3.98	3.89	4.07	3.98
M or F	2	M	3.62	3.99	4.08	3.90	3.85	4.11	4.04	4.00
M or F	2	F	3.79	4.19	4.41	4.13	4.02	3.91	3.97	3.97
M or F	1	M	3.96	4.12	4.09	4.06	4.10	4.12	4.21	4.15
M or F	1	F	3.86	4.45	4.10	4.14	3.85	4.22	3.97	4.01
M	1 or 2	M	3.90	4.21	4.13	4.08	4.02	4.26	4.03	4.11
M	1 or 2	F	3.95	4.56	4.64	4.38	4.02	4.31	4.32	4.22
F	1 or 2	M	3.68	3.90	4.04	3.87	3.93	3.97	4.22	4.04
F	1 or 2	F	3.71	4.08	3.87	3.89	3.85	3.82	3.62	3.76
M	1 or 2	M or F	3.92	4.38	4.39	4.23	4.02	4.29	4.18	4.16
F	1 or 2	M or F	3.69	3.99	3.95	3.88	3.89	3.90	3.92	3.90
M or F	1 or 2	M	3.79	4.06	4.09	3.98	3.98	4.12	4.13	4.07
M or F	1 or 2	F	3.83	4.32	4.26	4.14	3.94	4.06	3.97	3.99
M or F	2	M or F	3.71	4.09	4.25	4.02	3.94	4.01	4.01	3.98
M or F	1	M or F	3.91	4.28	4.09	4.10	3.98	4.17	4.09	4.08
M or F	1 or 2	M or F	3.81	4.19	4.17	4.06	3.96	4.09	4.05	4.03

iousness (Tables 5, 14), popularity, leadership (Tables 6, 15), and competitiveness (Tables 13, 22), and less high in severity of reaction to defeat (Tables 13, 22).

5. In the main, the mean ratings in leadership (Tables 6, 15), competitiveness (Tables 13, 22), and teasing (Tables 10, 19) were positively correlated with magnitude of difference (at least up to four years) in the age of the sibs.

6. Noteworthy is the greater hostility and competitiveness of first-born boys who have a sib two to four years younger, as compared to those with a sib less than two, or more than four years younger. The spacing-group differences in the following traits were, in most instances, significant: quarrelsomeness, revengefulness (Tables 8, 17), cruelty (Tables 9, 18), unkindness (Tables 7, 16), insistence on rights (Tables 12, 21), and tendency to tease (Tables 10, 19). In addition first-born boys with a sister, when the middle spacing is compared with the closest, were gauged the more jealous (Tables 11, 20), exhibitionistic (Tables 10, 19), and competitive (Tables 13, 22) but, as suggested above, also more friendly, gregarious (Tables 5, 14), popular, and better leaders (Tables 6, 15). Further, first-born boys with a sister two to four years younger were assessed as more jealous (Tables 11, 20), selfish (Tables 12, 21), revengeful (Tables 8, 17), insistent on their

rights (Tables 12, 21), and exhibitionistic (Tables 10, 19) than the parallel group at the widest spacing.

7. Consistent with the picture of male "improvement" in social attitude with spacing is the observation that the M2M [25-48]'s, as compared with the M2M [7-24]'s, scored, on the average, lower in quarrelsomeness, revengefulness (Tables 8, 17), cruelty (Tables 9, 18), and reaction to defeat (Tables 13, 22), but higher on exhibitionism (Tables 10, 19) and competitiveness (Tables 13, 22), while the M2M [49-72]'s, as compared with the M2M [7-24]'s, were judged less selfish (Tables 12, 21), less uncooperative (Tables 11, 20), less reactive to defeat (Tables 13, 22), but at the same time more competitive (Tables 13, 22) and given to teasing (Tables 10, 19).

8. Although the pattern in the traits investigated of the boy with an older sister altered little as the sib-age difference expanded, at the middle spacing, as compared with the closest, he was judged a less poor sport (Tables 13, 22) and less exhibitionistic (Tables 10, 19); and, when the former is compared with the boy whose sister was over four years older, he scored lower also on exhibitionism (Tables 10, 19) and tendency to tease (Tables 10, 19), as well as higher on

TABLE 11

MEANS OF THE NORMALIZED RATINGS FOR THE VARIOUS SUBGROUPS: JEALOUSY* AND UNCOOPERATIVENESS WITH PEERS

Description of Group			Jealousy*				Uncooperativeness with Peers			
Subject		Sibling	Age Difference Between Siblings in Months							
Sex	Ordinal Position	Sex	7–24	25–48	49–72	Total	7–24	25–48	49–72	Total
			Score Means				Score Means			
M	2	M	3.96	3.80	3.75	3.83	4.07	4.08	3.65	3.94
M	2	F	3.96	3.65	3.76	3.79	4.17	4.28	4.16	4.20
F	2	M	3.70	4.04	4.27	4.00	4.17	4.11	3.87	4.05
F	2	F	4.10	4.24	4.38	4.24	3.94	4.13	3.63	3.90
M	1	M	3.84	3.97	3.77	3.86	3.89	4.04	3.92	3.95
M	1	F	3.88	4.52	3.76	4.05	4.27	4.37	4.17	4.27
F	1	M	4.48	4.21	4.56	4.42	3.93	3.46	3.84	3.74
F	1	F	4.19	4.06	3.83	4.02	4.19	3.72	3.72	3.88
M	2	M or F	3.96	3.72	3.75	3.81	4.12	4.18	3.91	4.07
F	2	M or F	3.90	4.14	4.33	4.12	4.05	4.12	3.75	3.97
M	1	M or F	3.86	4.24	3.76	3.95	4.08	4.20	4.04	4.11
F	1	M or F	4.33	4.14	4.20	4.22	4.06	3.59	3.78	3.81
M or F	2	M	3.83	3.92	4.01	3.92	4.12	4.10	3.76	3.99
M or F	2	F	4.03	3.94	4.07	4.02	4.06	4.21	3.90	4.05
M or F	1	M	4.16	4.09	4.16	4.14	3.91	3.75	3.88	3.85
M or F	1	F	4.03	4.29	3.79	4.04	4.23	4.04	3.95	4.07
M	1 or 2	M	3.90	3.88	3.76	3.85	3.98	4.06	3.78	3.94
M	1 or 2	F	3.92	4.08	3.76	3.92	4.22	4.32	4.17	4.24
F	1 or 2	M	4.09	4.13	4.42	4.21	4.05	3.78	3.85	3.90
F	1 or 2	F	4.14	4.15	4.10	4.13	4.07	3.93	3.67	3.89
M	1 or 2	M or F	3.91	3.98	3.76	3.88	4.10	4.19	3.98	4.09
F	1 or 2	M or F	4.11	4.14	4.26	4.17	4.06	3.86	3.76	3.89
M or F	1 or 2	M	3.99	4.00	4.09	4.03	4.02	3.92	3.82	3.92
M or F	1 or 2	F	4.03	4.12	3.93	4.03	4.14	4.12	3.92	4.06
M or F	2	M or F	3.93	3.93	4.04	3.97	4.09	4.15	3.83	4.02
M or F	1	M or F	4.10	4.19	3.98	4.09	4.07	3.90	3.91	3.96
M or F	1 or 2	M or F	4.01	4.06	4.01	4.03	4.08	4.02	3.87	3.99

* Starred scale from reference 14, unstarred from reference 4.

TABLE 12

MEANS OF THE NORMALIZED RATINGS FOR THE VARIOUS SUBGROUPS: INSISTENCE ON RIGHTS AND SELFISHNESS

Description of Group			Insistence on Rights				Selfishness			
Subject		Sibling	Age Difference Between Siblings in Months							
Sex	Ordinal Position	Sex	7–24	25–48	49–72	Total	7–24	25–48	49–72	Total
			Score Means				Score Means			
M	2	M	4.18	3.93	4.21	4.10	4.16	4.18	3.61	3.99
M	2	F	3.78	4.13	3.64	3.85	4.16	4.19	4.16	4.17
F	2	M	3.54	3.67	3.87	3.69	3.91	4.32	3.84	4.02
F	2	F	4.21	3.84	3.95	4.00	3.45	4.19	3.93	3.86
M	1	M	3.59	4.87	4.15	4.08	3.71	4.06	3.86	3.88
M	1	F	4.18	4.60	4.05	4.28	4.09	4.40	3.83	4.11
F	1	M	4.22	4.08	4.56	4.29	3.95	3.91	4.14	4.00
F	1	F	3.69	3.71	4.07	3.82	4.02	4.00	4.09	4.04
M	2	M or F	3.98	4.03	3.92	3.98	4.16	4.19	3.89	4.08
F	2	M or F	3.88	3.76	3.91	3.85	3.68	4.26	3.89	3.94
M	1	M or F	3.89	4.54	4.10	4.18	3.90	4.23	3.85	3.99
F	1	M or F	3.95	3.90	4.31	4.05	3.98	3.96	4.12	4.02
M or F	2	M	3.86	3.80	4.04	3.90	4.04	4.25	3.73	4.00
M or F	2	F	4.00	3.99	3.79	3.93	3.80	4.19	4.05	4.01
M or F	1	M	3.90	4.28	4.35	4.18	3.83	3.98	4.00	3.94
M or F	1	F	3.93	4.16	4.06	4.05	4.05	4.20	3.96	4.07
M	1 or 2	M	3.89	4.21	4.18	4.09	3.94	4.12	3.74	3.93
M	1 or 2	F	3.98	4.37	3.84	4.06	4.12	4.29	4.00	4.14
F	1 or 2	M	3.88	3.88	4.21	3.99	3.93	4.11	3.99	4.01
F	1 or 2	F	3.95	3.78	4.01	3.91	3.73	4.10	4.01	3.95
M	1 or 2	M or F	3.93	4.29	4.01	4.08	4.03	4.21	3.87	4.03
F	1 or 2	M or F	3.92	3.83	4.11	3.95	3.83	4.11	4.00	3.98
M or F	1 or 2	M	3.88	4.04	4.20	4.04	3.93	4.12	3.86	3.97
M or F	1 or 2	F	3.96	4.07	3.93	3.99	3.93	4.20	4.00	4.04
M or F	2	M or F	3.93	3.89	3.92	3.91	3.92	4.22	3.89	4.01
M or F	1	M or F	3.92	4.22	4.21	4.12	3.94	4.09	3.98	4.00
M or F	1 or 2	M or F	3.92	4.06	4.06	4.01	3.93	4.16	3.93	4.01

kindness (Tables 7, 16) and insistence on rights (Tables 12, 21).

9. In contrast is the case of the girl with an older brother. Here the following traits were positively correlated with the degree of sib-age disparity: jealousy (Tables 11, 20), cruelty (Tables 9, 18), tendency to tease (Tables 10, 19), tendency to find fault (Tables 9, 18), and competitiveness (Tables 13, 22). An apparent peak was reached at the middle spacing in selfishness (Tables 12, 21) and quarrelsomeness (Tables 8, 17), while a dip occurred in intensity of reaction to defeat (Tables 13, 22).

10. The ratings given the girl with an older sister at the widest sib-age difference, as compared with the closest, were higher in the case of popularity, leadership (Tables 6, 15), and severity of reaction to defeat (Tables 13, 22). A formal peak in selfishness (Tables 12, 21), uncooperativeness (Tables 11, 20), and teasing (Tables 10, 19) was reached at the middle spacing, though the difference between the middle and widest spacing groups in these traits is, in most cases, insignificant. Popularity and kindness showed a dip at the two-to-four-year spacing.

11. With the expansion of the sib-age difference, girls with an older brother, in comparison with boys with an older brother, tended to shift from being the more friendly to peers, gregarious

TABLE 13

MEANS OF THE NORMALIZED RATINGS FOR THE VARIOUS SUBGROUPS: COMPETITIVENESS* AND BEHAVIOR REACTION TO DEFEAT IN COMPETITIVE ACTIVITY

Description of Group			Competitiveness*				Behavior Reaction to Defeat in Competitive Activity			
Subject		Sibling	Age Difference Between Siblings in Months							
Sex	Ordinal Position	Sex	7–24	25–48	49–72	Total	7–24	25–48	49–72	Total
			Score Means				Score Means			
M	2	M	3.65	4.14	4.08	3.95	4.52	3.86	3.98	4.12
M	2	F	3.72	3.62	3.45	3.60	4.21	3.69	3.97	3.96
F	2	M	3.84	4.43	4.24	4.17	3.98	3.28	4.41	3.89
F	2	F	3.88	3.65	3.92	3.82	3.69	3.67	4.14	3.84
M	1	M	3.72	3.92	4.30	3.98	4.42	4.08	3.84	4.11
M	1	F	3.64	4.35	4.31	4.10	4.14	4.05	3.85	4.01
F	1	M	4.26	4.40	4.22	4.29	4.17	3.96	4.10	4.08
F	1	F	3.79	3.86	4.07	3.91	4.03	4.15	4.08	4.08
M	2	M or F	3.68	3.88	3.76	3.78	4.37	3.77	3.97	4.04
F	2	M or F	3.86	4.04	4.08	3.99	3.84	3.48	4.28	3.86
M	1	M or F	3.68	4.13	4.31	4.04	4.28	4.07	3.84	4.06
F	1	M or F	4.02	4.13	4.14	4.10	4.10	4.06	4.09	4.08
M or F	2	M	3.74	4.28	4.16	4.06	4.25	3.57	4.19	4.00
M or F	2	F	3.80	3.64	3.68	3.71	3.95	3.68	4.05	3.90
M or F	1	M	3.99	4.16	4.26	4.14	4.30	4.02	3.97	4.10
M or F	1	F	3.71	4.10	4.19	4.00	4.08	4.10	3.96	4.05
M	1 or 2	M	3.68	4.03	4.19	3.97	4.47	3.97	3.91	4.12
M	1 or 2	F	3.68	3.99	3.88	3.85	4.17	3.87	3.91	3.98
F	1 or 2	M	4.05	4.41	4.23	4.23	4.08	3.62	4.25	3.98
F	1 or 2	F	3.83	3.76	3.99	3.86	3.86	3.91	4.11	3.96
M	1 or 2	M or F	3.68	4.01	4.03	3.91	4.32	3.92	3.91	4.05
F	1 or 2	M or F	3.94	4.08	4.11	4.05	3.97	3.77	4.18	3.97
M or F	1 or 2	M	3.87	4.22	4.21	4.10	4.27	3.80	4.08	4.05
M or F	1 or 2	F	3.76	3.87	3.94	3.86	4.02	3.89	4.01	3.97
M or F	2	M or F	3.77	3.96	3.92	3.88	4.10	3.62	4.12	3.95
M or F	1	M or F	3.85	4.13	4.23	4.07	4.19	4.06	3.96	4.07
M or F	1 or 2	M or F	3.81	4.05	4.07	3.98	4.14	3.84	4.04	4.01

* Starred scale from reference 14, unstarred from reference 4.

(Tables 5, 14) and popular (Tables 6, 15) to being the less friendly and gregarious. A similar shift between the sexes did not occur when the sib was a female.

12. When the sib-age difference was under two years, boys with an older brother showed no consistent differences from boys with an older sister, except in quarrelsomeness (Tables 8, 17), M2M's here exceeding M2F's. When the sib-age difference was over four years, the former group scored, in relation to the latter, higher in friendliness to peers, gregariousness (Tables 5, 14), cooperativeness (Tables 11, 20), popularity (Tables 6, 15), kindness (Tables 7, 16), competitiveness (Tables 13 22), insistence on rights (Tables 12, 21), and leadership (Tables 6, 15), as well as lower in selfishness (Tables 12, 21) and tendency to tease (Tables 10, 19).

13. Only at the two-to-four year spacing were the mean ratings for boys for insistence on rights consistently higher than for girls (Tables 12, 21). At the closest spacing, however, F1M's exceeded significantly M1M's.

14. When the sibs differed in age by less than two years, no significant group differences in exhibitionism (Tables 10, 19) were apparent; but when the age difference was over four years, those from opposite-sex pairs tended to be

TABLE 14

Analysis of the Variance for Various Traits

Source of Variance	*df*	Trait					
		Friendliness to Peers			Gregariousness		
		Sum of Squares	Estimate of Variance	*F*	Sum of Squares	Estimate of Variance	*F*
Total	383	326.6677			309.8936		
Between groups	23	43.4475	1.8892		31.7679	1.3812	
Within groups	360	283.2202	.7867		278.1257	.7726	
Between replications	15	15.8619	1.0575		5.7398	.3827	
Residual	345	267.3583	.7750		272.3859	.7895	
Between Variables							
Between sex groups	1	8.5682	8.5682	11.1	3.5075	3.5075	4.4
Between sibling's-sex groups	1	1.5606	1.5606		2.0097	2.0097	
Between ordinal-position groups	1	1.5939	1.5939		1.7713	1.7713	
Between spacing groups	2	7.5694	3.7847	4.9	4.8093	2.4046	3.0
First-order interactions							
Subject's sex—sibling's sex	1	.0048	.0048		.0284	.0284	
Subject's sex—ordinal position	1	.4121	.4121		1.0458	1.0458	
Subject's sex—spacing	2	7.6317	3.8159	4.9	6.3172	3.1586	4.0
Sibling's sex—ordinal position	1	.0341	.0341		.9087	.9087	
Sibling's sex—spacing	2	1.1768	.5884		.0209	.0104	
Ordinal position—spacing	2	.9123	.4562		1.0686	.5343	
Second-order interactions							
Subject's sex—sibling's sex—ordinal position	1	.0018	.0018		2.2817	2.2817	
Subject's sex—sibling's sex—spacing	2	3.5544	1.7772		1.7022	.8511	
Subject's sex—ordinal position spacing	2	1.5089	.7544		2.3573	1.1786	
Sibling's sex—ordinal position spacing	2	4.9809	2.4904	3.2	2.3235	1.1618	
Third-order interaction							
Subject's sex—sibling's sex—ordinal position—spacing	2	3.5566	1.7783		1.6158	.8079	

gauged higher on the trait than those from same-sex pairs. The same trend obtained at the middle spacing among first-borns, but the M2M [25-48]'s exceeded the M2F [25-48]'s.

15. Only at the middle spacing were second-born girls with a brother, as compared to those with a sister, assessed as the more friendly (Tables 5, 14), popular (Tables 6, 15), kind (Tables 7, 16), competitive (Tables 13, 22), better leaders (Tables 6, 15), and less critical (Tables 9, 18).

16. When the sib-age disparity was over four years, those from sib pairs similar in sex tended to be rated more popular (Tables 6, 15) than those from opposite-sex pairs; but at the middle spacing M1F's exceeded M1M's, and at both the under-two and two-to-four-year spacing F2M's were rated higher than F2F's.

17. Among second-borns sex differences in jealousy (Tables 11, 20) increased with spacing, being insignificant at the closest and significant at the widest, girls at the latter interval surpassing boys.

18. At the under-two and over-four-year spacings F1M's scored higher in quarrelsomeness (Tables 8, 17), insistence on rights (Tables 12, 21), jealousy (Tables 11, 20), and criticalness (Tables 9, 18) than did M1M's.

19. At the closest and widest spacings second-born boys received a higher mean rating on cruelty (Tables 9, 18) than did girls, but no significant sex difference obtained among first-borns.

20. When the sib-age difference was less than two years, and also among first-borns at the middle spacing, boys were rated higher in revenge-

TABLE 15

ANALYSIS OF THE VARIANCE FOR VARIOUS TRAITS

Source of Variance	df	Trait					
		Popularity			Leadership		
		Sum of Squares	Estimate of Variance	F	Sum of Squares	Estimate of Variance	F
Total	383	319.4908	.8342		331.2067	.8648	
Between groups	23	27.6633	1.2028		34.6563	1.5068	
Within groups	360	291.8275	.8106		296.5504	.8238	
Between replications	15	12.8728	.8582		16.8014	1.1201	
Residual	345	278.9547	.8086		279.7490	.8109	
Between variables							
Between sex groups	1	4.8196	4.8196	6.0	2.9156	2.9156	3.6
Between sibling's-sex groups	1	1.6511	1.6511		3.6895	3.6895	4.6
Between ordinal-position groups	1	1.0375	1.0375		.7298	.7298	
Between spacing groups	2	3.4951	1.7476		7.5748	3.7874	4.7
First-order interactions							
Subject's sex—sibling's sex	1	.0585	.0585		1.9580	1.9580	
Subject's sex—ordinal position	1	1.7931	1.7931		.0198	.0198	
Subject's sex—spacing	2	1.2559	.6280		2.3950	1.1975	
Sibling's sex—ordinal position	1	1.9551	1.9551		.3788	.3788	
Sibling's sex—spacing	2	.0999	.0500		1.5514	.7757	
Ordinal position—spacing	2	1.2066	.6033		.5322	.2661	
Second-order interactions							
Subject's sex—sibling's sex—ordinal position	1	.0096	.0096		3.9204	3.9204	4.8
Subject's sex—sibling's sex—spacing	2	5.5654	2.7827	3.4	4.2219	2.1110	
Subject's sex—ordinal position—spacing	2	1.0055	.5028		1.6235	.8118	
Sibling's sex—ordinal position—spacing	2	2.8043	1.4022		2.7947	1.3974	
Third-order interaction							
Subject's sex—sibling's sex—ordinal position—spacing	2	.9061	.4530		.3509	.1755	

fulness (Tables 8, 17) than girls, but no significant sex difference was apparent among other groups.

21. When the sib spacing was close, second-born girls received lower ratings on severity of reaction to defeat (Tables 13, 22) than when the sib spacing was more than four years. The reverse trend seemed to hold for second-born boys.

Sib's-Sex Group Differences[3,4]

22. The variable, sex of the sibling, was most frequently involved in interactions with the variable, child's sex.

23. In the main, children with a brother received higher ratings, on the average, in competitiveness and leadership than did those with a sister. However M1F [25-48]'s were judged more competitive and better leaders than M1M [25-48]'s.

24. In fact, first-born boys with a sister, when the age difference was two to four years, exceeded boys with a brother in ratings on all the "positive" social traits, as well as in jealousy, exhibitionism, and competitiveness. Similarly, second-born girls with a brother at the middle spacing tended to be assessed higher than second-born girls with a sister in not only all "positive" social traits but also in competitiveness; and the

---

[3] From this point on no table references will be given in the outline, as the reader doubtless by this time knows his way around among the data.

[4] See also previous section, Items 3, 6, 9, 10, 11, 13, 14, 15, 16, 18 and 19.

TABLE 16

Analysis of the Variance for Various Traits

Source of Variance	df	Trait: Kindness — Sum of Squares	Estimate of Variance	F
Total	383	312.7300		
Between groups	23	23.1095	1.0048	
Within groups	360	289.6205	.8045	
Between replications	15	21.2527	1.4168	
Residual	345	268.3678	.7779	
Between variables				
Between sex groups	1	1.8068	1.8068	
Between sibling's-sex groups	1	.1434	.1434	
Between ordinal-position groups	1	.1094	.1094	
Between spacing groups	2	1.6070	.8035	
First-order interactions				
Subject's sex—sibling's sex	1	.9984	.9984	
Subject's sex—ordinal position	1	.0451	.0451	
Subject's sex—spacing	2	1.0089	.5044	
Sibling's sex—ordinal position	1	.0542	.0542	
Sibling's sex—spacing	2	3.6232	1.8116	
Ordinal position—spacing	2	1.5021	.7511	
Second-order interactions				
Subject's sex—sibling's sex—ordinal position	1	.2109	.2109	
Subject's sex—sibling's sex—spacing	2	4.8473	2.4236	3.1
Subject's sex—ordinal position—spacing	2	3.6531	1.8266	
Sibling's sex—ordinal position—spacing	2	2.0393	1.0196	
Third-order interaction				
Subject's sex—sibling's sex—ordinal position—spacing	2	1.4604	.7302	

former tended also to be rated lower in criticalness and not quite significantly lower in severity of reaction to defeat. At the closest and widest spacings few parallel sib-sex group differences were significant.

25. First-borns with a sib of opposite sex were gauged more quarrelsome, jealous, fault-finding, exhibitionistic, insistent on rights than those with a sib of the same sex. No similar trend was observed in the case of second-borns. If anything, the trend was the reverse. M2M [7-24]'s were judged more quarrelsome than M2F [7-24]'s; M2M [25-48]'s, more exhibitionistic than M2F [25-48]'s; F2F [25-48]'s, more fault-finding than F2M [25-48]'s, and M2M [49-72]'s than M2F [49-72]'s.

26. In friendliness to peers M1M [7-24]'s exceeded M1F [7-24]'s; and M2M [49-72]'s, M2F [49-72]'s; but, as stated above, M1F [25-48]'s were gauged higher than M1M [25-48]'s.

27. Boys with a much older brother received a higher assessment, on the average, than boys with a much older sister in not only friendliness to peers, but in gregariousness, leadership, popularity, kindness, competitiveness, insistence on rights, cooperativeness, and were judged also less selfish as well as less inclined to tease.

Ordinal-Position Group Differences[5]

28. Among the interactions involving the ordinal-position variable, those with the variable, child's sex, were the most frequent. The frequency of significant interactions suggests the need for caution in generalization about the effects of birth-order.

29. There were few ordinal-position group differences when the child and his sib were similar in sex (see especially girls with a sister) but, when opposite in sex, many differences were noted.

---

[5] See also previous sections, Items 2, 6, 14, 16, 19, 21, 23, 24, 25, 26.

30. First-borns tended to be rated, on the average, as more competitive than second-borns, as Adler (1) and Sanford *et al.* (16) have also suggested, the ordinal-position difference increasing with spacing.

31. First-born boys from opposite-sex sibling pairs, in comparison with parallel groups of second-borns, tended at the two wider spacings to be judged more aggressive socially—i.e., to be more friendly to peers, gregarious, and better leaders, while at the same time, more insistent on their rights and competitive; and, furthermore, at the middle spacing, more exhibitionistic, jealous, quarrelsome, and revengeful also. A somewhat similar trend is to be noted for girls in that girls with a brother their junior, in comparison with girls with a brother their senior, were gauged at one or another spacing more quarrelsome, cruel, insistent on rights, jealous, gregarious, competitive, given to teasing, and exhibitionistic. Differences, however, tended to be most marked at the close spacing.

32. Those with a sib two to four years younger were evaluated as more upset by defeat, on the average, than those with a sib correspondingly older, and, if the child was a male, more revengeful also.

33. Second-borns from same-sex pairs at the sib spacing under two years scored higher, on the average, on the trait, insistence on rights, than first-borns; whereas in the case of those from opposite-sex pairs, the situation was reversed. At the spacings over two years first-borns tended to be rated more insistent.

34. Boys with an older brother were assessed as more hostile, in the main, than those with

TABLE 17

ANALYSIS OF THE VARIANCE FOR VARIOUS TRAITS

Source of Variance	*df*	Trait					
		Quarrelsomeness			Revengefulness		
		Sum of Squares	Estimate of Variance	*F*	Sum of Squares	Estimate of Variance	*F*
Total	383	345.3629			350.5617	.9153	
Between groups	23	37.0519	1.6110		20.6773	.8990	
Within groups	360	308.3110	.8564		329.8844	.9163	
Between replications	15	19.6742	1.3116		24.0277	1.6018	
Residual	345	288.6368	.8366		305.8567	.8865	
Between variables							
Between sex groups	1	9.8785	9.8785	11.8	8.6280	8.6280	9.7
Between sibling's-sex groups	1	.4036	.4036		.1785	.1785	
Between ordinal-position groups	1	.6675	.6675		.0353	.0353	
Between spacing groups	2	1.2759	.6380		.9653	.4826	
First-order interactions							
Subject's sex—sibling's sex	1	1.0220	1.0220		.0380	.0380	
Subject's sex—ordinal position	1	3.2801	3.2801	3.9	.0944	.0944	
Subject's sex—spacing	2	1.6714	.8357		1.8394	.9197	
Sibling's sex—ordinal position	1	.0087	.0087		.2699	.2699	
Sibling's sex—spacing	2	.0471	.0236		.0976	.0488	
Ordinal position—spacing	2	.6797	.3399		.9899	.4950	
Second-order interactions							
Subject's sex—sibling's sex—ordinal position	1	6.5130	6.5130	7.8	.0376	.0376	
Subject's sex—sibling's sex—spacing	2	.4294	.2147		.6030	.3016	
Subject's sex—ordinal position—spacing	2	9.5303	4.7652	5.7	6.6354	3.3177	3.7
Sibling's sex—ordinal position—spacing	2	.0218	.0109		.1085	.0542	
Third-order interaction							
Subject's sex—sibling's sex—ordinal position—spacing	2	1.6229	.8114		.1565	.0782	

TABLE 18

ANALYSIS OF THE VARIANCE FOR VARIOUS TRAITS

Source of Variance	*df*	Trait					
		Cruelty			Tendency to Find Fault		
		Sum of Squares	Estimate of Variance	*F*	Sum of Squares	Estimate of Variance	*F*
Total	383	273.7443	.7147		340.8929		
Between groups	23	22.1043	.9611		19.2066	.8351	
Within groups	360	251.6400	.6990		321.6863	.8936	
Between replications	15	18.7210	1.2481		18.8026	1.2535	
Residual	345	232.9190	.6751		302.8837	.8779	
Between variables							
Between sex groups	1	5.6042	5.6042	8.3	.0030	.0030	
Between sibling's-sex groups	1	.0488	.0488		.0449	.0449	
Between ordinal-position groups	1	.4490	.4490		.5528	.5528	
Between spacing groups	2	.5325	.2663		3.9914	1.9957	
First-order interactions							
Subject's sex—sibling's sex	1	.3985	.3985		.3272	.3272	
Subject's sex—ordinal position	1	1.3908	1.3908		.0022	.0022	
Subject's sex—spacing	2	.0154	.0077		.9003	.4502	
Sibling's sex—ordinal position	1	.1722	.1722		.0146	.0146	
Sibling's sex—spacing	2	.0228	.0114		3.5460	1.7730	
Ordinal position—spacing	2	3.3715	1.6858		1.0166	.5083	
Second-order interactions							
Subject's sex—sibling's sex—ordinal position	1	.2227	.2227		5.1038	5.1038	5.8
Subject's sex—sibling's sex—spacing	2	.2048	.1024		2.1326	1.0663	
Subject's sex—ordinal position—spacing	2	7.3787	3.6894	5.5	.0692	.0346	
Sibling's sex—ordinal position—spacing	2	1.2802	.6401		.0651	.0326	
Third-order interaction							
Subject's sex—sibling's sex—ordinal position—spacing	2	1.0122	.5061		1.4369	.7185	

a younger brother, except at the middle spacing, where the relation was the reverse.

35. While at the close spacing the boy with a younger sister was judged more quarrelsome, if anything, than the boy with a brother his junior, the boy with an older sister was assessed less quarrelsome than the boy with a brother his senior.

36. First-born girls were appraised more cruel than second-borns when the sibs differed by less than two years in age, whereas among boys the ordinal-position group's relationship, though insignificant, was the reverse.

Sex-Group Differences[6]

37. The variable, child's sex, was involved the most frequently in interactions with our other independent variables and about equally frequently with each of the three. More of the traits studied were related to child's sex in some way than was the case with any of the other variables.

38. Only in uncooperativeness with peers and tendency to tease was there any simple sex difference to be noted, boys in these traits being rated higher than girls. The finding relative to teasing agrees with that of Berne (3).

39. Girls received a higher mean rating in jealousy than did boys. [This agrees with the findings of Ross (15), Sewall (19) and Foster (7).] The most consistent differences are to be noted when first-born girls with a brother are

---

[6] See also previous sections, Items 4, 8, 9, 11, 13, 17, 18, 19, 20, 21, 22, 25, 32.

compared with first-born boys with a brother or when children who are in opposite-sex sib pairs are compared. Among second-borns, sex difference in jealousy tended to widen with spacing, being insignificant at the close spacing and large at the widest.

40. Girls, on the average, were scored consistently as more friendly, gregarious, popular, better leaders, and less reactive to defeat than boys only under the condition in which the sib differed in age by under two years.

41. In the case of boys and the traits named in 40, the assessment means tended to increase with spacing, while no similar relationship to spacing was apparent in the case of girls.

42. Boys were gauged generally more overtly hostile than girls, but first-born girls with a brother at the close spacing received a higher mean rating in quarrelsomeness, criticalness, insistence on rights, jealousy, and competitiveness than first-born boys with a brother, while at the same time tending to score higher in all the socially "positive" traits.

43. The assessment in revengefulness, uncooperativeness, selfishness, tendency to tease, cruelty, exhibitionism, and quarrelsomeness of second-born boys with a sister was higher than that for second-born girls with a sister, only at the closest and widest spacings.

44. Boys were judged, on the average, as more revengeful than girls when the sib spacing was close; but when the sib-age disparity was two to four years, a sex difference obtained only in the case of first-borns. At the widest spacing there were no consistent sex differences in revengefulness apparent.

TABLE 19

ANALYSIS OF THE VARIANCE FOR VARIOUS TRAITS

Source of Variance	*df*	Trait					
		Tendency to Tease			Exhibitionism		
		Sum of Squares	Estimate of Variance	*F*	Sum of Squares	Estimate of Variance	*F*
Total	383	323.6876			343.6622		
Between groups	23	32.9670	1.4333		29.6528	1.2893	
Within groups	360	290.7206	.8076		314.0094	.8722	
Between replications	15	20.5192	1.3679		14.1061	.9404	
Residual	345	270.2014	.7832		299.9033	.8693	
Between variables							
Between sex groups	1	9.8528	9.8528	12.6	6.2961	6.2961	7.2
Between sibling's-sex groups	1	1.1019	1.1019		.9411	.9411	
Between ordinal-position groups	1	1.6682	1.6682		.6460	.6460	
Between spacing groups	2	9.9996	4.9998	6.4	.9242	.4621	
First-order interactions							
Subject's sex—sibling's sex	1	1.2004	1.2004		3.5593	3.5593	4.1
Subject's sex—ordinal position	1	.0440	.0440		.3261	.3261	
Subject's sex—spacing	2	.8389	.4194		.9653	.4826	
Sibling's sex—ordinal position	1	.0622	.0622		.3882	.3882	
Sibling's sex—spacing	2	1.2097	.6048		.2205	.1102	
Ordinal position—spacing	2	.8657	.4328		.1044	.0522	
Second-order interactions							
Subject's sex—sibling's sex—ordinal position	1	.1773	.1773		2.2956	2.2956	2.6
Subject's sex—sibling's sex—spacing	2	.5484	.2742		3.1550	1.5775	
Subject's sex—ordinal position—spacing	2	4.3909	2.1954		2.4463	1.2232	
Sibling's sex—ordinal position—spacing	2	.6785	.3392		1.6095	.8048	
Third-order interaction							
Subject's sex—sibling's sex—ordinal position—spacing	2	.3285	.1642		5.7752	2.9876	3.3

TABLE 20

ANALYSIS OF THE VARIANCE FOR VARIOUS TRAITS

Source of Variance	*df*	Trait					
		Jealousy			Uncooperativeness with Group		
		Sum of Squares	Estimate of Variance	*F*	Sum of Squares	Estimate of Variance	*F*
Total	383	314.5296			339.1065		
Between groups	23	27.4597	1.1939		19.7481	.8586	
Within groups	360	287.0699	.7974		319.3584	.8871	
Between replications	15	19.0831	1.2722		10.5938	.7062	
Residual	345	267.9868	.7768		308.7646	.8950	
Between variables							
Between sex groups	1	7.6671	7.6671	9.9	3.6797	3.6797	4.1
Between sibling's-sex groups	1	.0119	.0119		1.9281	1.9281	
Between ordinal-position groups	1	1.2060	1.2060		.3946	.3946	
Between spacing groups	2	.1147	.0574		2.9753	1.4877	
First-order interactions							
Subject's sex—sibling's sex	1	.6468	.6468		2.1108	2.1108	2.4
Subject's sex—ordinal position	1	.0699	.0699		.9471	.9471	
Subject's sex—spacing	2	2.4071	1.2036		1.3622	.6811	
Sibling's sex—ordinal position	1	1.0859	1.0859		.6575	.6575	
Sibling's sex—spacing	2	1.0444	.5222		.1518	.0759	
Ordinal position—spacing	2	1.5391	.7696		1.9604	.9802	
Second-order interactions							
Subject's sex—sibling's sex—ordinal position	1	4.7259	4.7259	6.1	.3462	.3462	
Subject's sex—sibling's sex—spacing	2	.5165	.2582		.9037	.4514	
Subject's sex—ordinal position—spacing	2	4.8882	2.4441	3.1	1.4735	.7368	
Sibling's sex—ordinal position—spacing	2	1.3685	.6842		.7939	.3970	
Third-order interaction							
Subject's sex—sibling's sex—ordinal position—spacing	2	.1677	.0838		.0633	.0316	

Obverse Groups

Since we shall have theories relative to the influence of sibs on each other, it will be of interest to look at the relation of the obverse groups. From them we can get some idea of what the sib of any given type of child is like. For example, we may infer, though the factor of age is not controlled, what the sister of the boy with an older sister is like by looking at the F1M group—i.e., the girls with a younger brother (see Tables 5-13).

45. Boys were given higher ratings than girls in cruelty at the middle spacing among first-borns and at the close and wide spacings among second-borns.

46. At the under-two and over-four-year spacings M1M's received a lower evaluation in quarrelsomeness and insistence on rights than F1M's, but M2M's were judged to exceed F2M's in these traits.

47. It is noteworthy that the F1M group was rated highest, on the average, on all the "positive" social traits, while the M2F, the obverse group, was rated lowest, on the average (see Tables 5-7). A similar strong consistent negative relation was not observed in the other obversely related pairs, though a hint can be seen. If we rank the totals for the first- and second-borns and correlate the ranks of the obverse groups, the rho's in the case of most of

the "positive" social traits are negative, though of course the rho's based on four cases are insignificant. The rho's for the totals are —.80, —1.00, —.80, —.80, and —.20, respectively, for friendliness to peers, gregariousness, popularity, leadership, and kindness. The rho's for the "negative" traits are not consistently negative and practically all are small. The negative relations, however, may be merely a reflection of the influence of the sex variable. If the trends noted in the case of the "positive" traits can be trusted, it would not be too difficult to understand some of the items. If one sib were high in leadership, for instance, then the other might get into the habit of following. The negative sib relations in the case of friendliness and gregariousness are not so easy to conceptualize and may well be chiefly an expression of the sex variable. The relations hint possibly of the operation, in part, of a compensatory mechanism, though they could be understood, as in the case of leadership, in terms of dependence. We should in any case expect the sign of the correlations to vary with the traits. We shall comment in the discussion on a few other obverse group relationships. There are some interesting spacing-pattern similarities.

III. Discussion

All of the details of the relationships just enumerated we shall not attempt to account for, but a dozen or two demand comment. Let us say parenthetically that, although the hypotheses we shall

TABLE 21

Analysis of the Variance for Various Traits

Source of Variance	*df*	Trait					
		Insistence on Rights			Selfishness		
		Sum of Squares	Estimate of Variance	*F*	Sum of Squares	Estimate of Variance	*F*
Total	383	356.6076			333.4160		
Between groups	23	33.0254	1.4359		17.2167	.7486	
Within groups	360	323.5822	.8988		316.1993	.8783	
Between replications	15	16.1877	1.0792		12.9230	.8615	
Residual	345	307.3945	.8910		303.2763	.8791	
Between variables							
Between sex groups	1	1.6656	1.6656		.1971	.1971	
Between sibling's-sex groups	1	.3857	.3857		.4902	.4902	
Between ordinal-position groups	1	3.5401	3.5401	4.0	.0022	.0022	
Between spacing groups	2	1.3888	.6944		4.2976	2.1488	
First-order interactions							
Subject's sex—sibling's sex	1	.0825	.0825		1.5075	1.5075	
Subject's sex—ordinal position	1	.0001	.0001		.7794	.7794	
Subject's sex—spacing	2	5.8081	2.9040	3.2	1.8143	.9072	
Sibling's sex—ordinal position	1	.7912	.7912		.3851	.3851	
Sibling's sex—spacing	2	2.1520	1.0760		.3303	.1652	
Ordinal position—spacing	2	1.9410	.9705		.8335	.4168	
Second-order interactions							
Subject's sex—sibling's sex—ordinal position	1	9.2287	9.2287	10.4	.2350	.2350	
Subject's sex—sibling's sex—spacing	2	.6916	.3458		.2587	.1294	
Subject's sex—ordinal position—spacing	2	1.9031	.9515		2.9616	1.4808	
Sibling's sex—ordinal position—spacing	2	.5235	.2618		2.9218	1.4609	
Third-order interaction							
Subject's sex—sibling's sex—ordinal position—spacing	2	2.9234	1.4617		.2024	.1012	

TABLE 22

ANALYSIS OF THE VARIANCE FOR VARIOUS TRAITS

Source of Variance	*df*	Trait: Competitiveness: Sum of Squares	Competitiveness: Estimate of Variance	Competitiveness: *F*	Behavior Reaction to Defeat in Competitive Activity: Sum of Squares	Behavior Reaction: Estimate of Variance	Behavior Reaction: *F*
Total	383	322.2595			339.6890	.8869	
Between groups	23	30.5315	1.3275		26.3502	1.1457	
Within groups	360	291.7280	.8104		313.3388	.8704	
Between replications	15	22.9325	1.5288		18.1761	1.2117	
Residual	345	268.7955	.7791		295.1627	.8555	
Between variables							
Between sex groups	1	1.9267	1.9267		.6551	.6551	
Between sibling's-sex groups	1	5.8856	5.8856	7.6	.6419	.6419	
Between ordinal-position groups	1	3.1683	3.1683	4.1	1.3301	1.3301	
Between spacing groups	2	5.2107	2.6050	3.3	6.3562	3.1781	3.7
First-order interactions							
Subject's sex—sibling's sex	1	1.4235	1.4235		.2370	.2370	
Subject's sex—ordinal position	1	.5430	.5430		.8103	.8103	
Subject's sex—spacing	2	.6670	.3335		6.6127	3.3064	3.9
Sibling's sex—ordinal position	1	1.0987	1.0987		.0651	.0651	
Sibling's sex—spacing	2	1.0352	.5176		1.8181	.9091	
Ordinal position—spacing	2	.8367	.4184		5.4384	2.7192	3.2
Second-order interactions							
Subject's sex—sibling's sex—ordinal position	1	1.4627	1.4627		.0030	.0030	
Subject's sex—sibling's sex—spacing	2	1.7615	.8808		1.0135	.5068	
Subject's sex—ordinal position—spacing	2	1.6745	.8372		.7505	.3752	
Sibling's sex—ordinal position—spacing	2	3.7383	1.8692		.1638	.0819	
Third-order interaction							
Subject's sex—sibling's sex—ordinal position—spacing	2	.0991	.0496		.4545	.2272	

now offer have an *ad hoc* character, we shall bring to bear upon the issues some of the data from other areas of our research. The later data are not all presented here in detail because they will be, or have already been, published elsewhere (10, 11).

The ratings bearing on the "positive" social attitudes just described indicate there is a substantial degree of correlation among them. This positive relationship could, of course, be viewed as a halo effect, or a lack of discrimination on the part of the judges, though it seems reasonable to assume the traits would be correlated to some extent because of psychological mechanisms in the subject rather than merely in the judge. The traits were, in fact, chosen for study partly because of their putative overlap. We wanted the traits investigated to be rather specifically defined, yet we also hoped to cover such significant attitude *areas* as hostility, dependence, and social interest. Hostility, for instance, expresses itself in many ways. We wondered, for instance, whether a child who is consistently overpowered by an older sib would develop the habit of resorting to the more indirect expressions of hostility. That a halo effect is not wholly responsible for the variable correlations is suggested by the fact that the detail of the relations of the various traits in our independent variables shows variations; and in some instances, both "positive" and "negative" traits are similarly related to the sibling variables rather than in opposite ways, as the halo effect would presumably dictate.

Sib spacing. The threefold fact that boys, with the exception of the ones with an older sister, tended to become more sociable and socially dynamic, i.e., to be better leaders and more friendly to peers, more popular, and more competitive, as the disparity in age between them and their sibs increased, while girls, in the main, did not show this trend, is interesting (Items 4, 7). We must remind the reader, however, that the higher ratings at the wider spacings in social interest and effectiveness with peers are often paralleled by higher ratings on a number of attitudes or behaviors reflecting hostility (Item 6). It is also worthy of note that it is the obversely related groups—F1F and F2F, and F1M and M2F—which do not reflect much in the way of spacing differences in the so-called "positive" traits (Item 2; Tables 5-7).

We suspect that a child's reaching out to peers may be a function in part of his involvement with, or the degree of satisfyingness of, his sib. When the age difference between sibs is small, we think they tend to be more involved with each other—not only in a rivalrous way but also in real companionship and dependence. Interests overlap greatly and companionship may even be sufficiently satisfying so that contact with other children is not sought. Relevant to this general issue is the fact that, when we asked the children in the interview we had with each, whether they thought they would be happier without the sib, fewer second-borns in the close spacing group, as compared with those in the wide spacing groups, wished to dispense with the sib or preferred others' company. When the children whose sibs differed from them little in age were asked why they felt as they did, their response tended to revolve around the idea that they would be lonely without the sib. First-borns generally *less* frequently than second-borns thought they would like to be rid of the sib. They seemed, however, more involved with or focused on their relation to their parents, whereas the second-borns appeared to focus on the sib relatively more than did the first-borns. In other words, our subjects seemed more concerned about those above them than with those below them in the age hierarchy. For instance, second-borns more frequently than first told us they preferred to play with their sibs than to play with other children. However, significantly more frequently than first-borns the former reported that they rarely played with their sibs. This is probably correct, for, if the sib is younger, he will be at home most of the day, whereas, if older, he will be at school much of the time. It is also congruent that our first-borns, especially at the spacings over two years, dealt in the CAT (Children's Apperception Test) stories (2) they told, consistently less frequently with sibling activities and sibling interaction themes than did second-borns and the mention of sib cooperative activities decreased with spacing.

If a child and his sib differ in age by more than two years, then the former, we suspect, by virtue (among other things) of having a larger coterie of friends of his own, feels not quite so dependent on the latter or even on other family members, as would be the case were the sib age difference small. (Relevant to Items 4, 5, and 6.) Our subjects reported *decreasingly* with the widening of the age gap between them and their sibs that they played much with the latter and the latters' friends. It is significant for our hypothesis, too, that the percentage of children, male and female, who expressed in the interview an indifference as to sex of playmate diminished the more unlike in age the sibs were; and the percentage who indicated they preferred playmates of their own sex waxed. The most significant alteration in pattern occurred between the close and middle spacing (see Items 5, 7, 8, 9, and 10). It is striking that the preference frequency among boys with a slightly older sister for female playmates was somewhat higher than their preference for males (42 vs. 33 per cent), but when the sister was two to four years older, the corresponding per cents were 7 and 80. Having more companions of his own sex, age, and choosing (especially true of second-borns) is probably very important for the child, for then he is more evenly matched with his associates and has more opportunity to develop self-confidence in his social relations with the latter. If a second-born, the child whose sib differs much in age would be less a mere hanger-on than, for instance, one whose sib is two to four years older. It is significant in this connection that our data show leadership and competitiveness to be positively correlated with sib spacing (Item 5). The following facts are still further evidence of the lesser influence of the sib (especially in the case of the second-born) on the character of the playmates of the child as the sib age disparity becomes greater. Second-borns, except F2M's, as the gap in sib age widened, more frequently stated in the interview that their best friend was of their own sex and, with declining frequency named as playmates children who were older than they. Congruent with the hypotheses that, as the sib-age difference expands, the child tends to have more playmates of his own sex, age, and choosing (more marked among second-borns) and that this condition is favorable to the development of self-confidence as well as learning to be effective with his peers, is the observation that boys in the wider spacing groups, in contrast to those in the close, were judged by their teachers to be more self-confident, active, aggressive, ambitious, better sports, possessed of a wider range of interests, and to be less apprehensive socially (data not yet published).

The greater self-confidence, peer interest, and friendliness in first-born boys which occurred in the groups with the wider sib-age disparity (Items 4, 5, 6), we think may be due in part also to lesser stress from the displacement experience. The first-born boy whose sib arrives after he is four or more years old, for instance, has had his mother's attention uncontested for a considerable time. The displacement experience comes to him after he feels relatively more sure of his relationship with his mother, as well as after more mature social interests have developed. The necessary attention the mother gives the infant sib, for instance, which, if the spacing is wide, is what the sib would be at the time of our contact with the children, may also seem more reasonable and less discriminating to a child of five or six than to a child of two. It is interesting that in the CAT stories of our first-borns the parent-favoritism theme decreased the greater the difference in age between child and sib. (The incidence of these themes was too low for the difference to be significant, however.) As has been said, the more expansive social adjustment of the children whose sibs differ much from them in age may be due in part to the greater association of the child with peers of his own age, sex, and choosing. It is worthy of note in this connection that in the CAT stories of the M2M's, for instance, there were fewer mentions of sibling activities by the children whose sibs differed from them most in age, also more cases in which no sib was even mentioned in the stories. There were fewer deception themes, too, and fewer mentions of moral issues—of right and wrong.

Still other forces may be at work to bring about a correlation of sib spacing interval and favorable attitudes toward peers on the part of boys with an older brother (Items 4, 7, 27). Since our subjects came only from two-child families, the second child then retained his position of youngest. The older the sib relatively, the more the second child–or at least the boy–is likely to be indulged (see 18). He will decreasingly be expected to fight out his difficulties with the older sib and others, and will have more protection; he will have more associates at his own level; and he will have relatively more of his mother's attention uncontested, because his sib is off at school much of the time. His mother will probably be more relaxed. All of these factors should reduce the child's dependency drive [i.e., if frustration intensifies drive, as Whiting believes (22)] and make him at the wide spacing relatively more outgoing and friendly. It is also very significant, we think, that the group type of the older brother (see M1M) was rated higher in friendliness, gregariousness (Tables 5, 14), popularity, and leadership (Tables 6, 15) when the spacing was wide then when close (Item 4). This likely trend, then, in the older sib we should expect to be paralleled by less threat to the younger and, hence, by correspondingly more effective social attitudes on the part of the latter. Since the sibs' interests overlap less when the spacing is wide, the older child can afford to be more protective of, and less competitive with, the younger.

If, on the other hand, the boy's sib is an older sister, instead of an older brother, a somewhat different weight of forces may obtain (see Items 2, 8, 12, 27). With her and her friends the boy will be encouraged to associate little, when the spacing is wide. Forty-seven per cent of the M2F [49-72]'s, for instance, reported in the interview we had that they *rarely* play with the sib, whereas 13 per cent of the M2M [49-72]'s gave this estimate. Also with a high frequency the M2F [49-72]'s confessed they were badly treated by the sib's friends. We believe, in addition, that the older sister is likely to badger and dominate her younger brother because she is very jealous (see F1M), thus causing the mechanisms making for retreat and dependency in the latter to be emphasized (1). M2F [49-72]'s, as compared with M2M [49-72]'s, the reader should note, were rated less friendly to peers, gregarious, popular, kind, competitive, cooperative, insistent on rights, poorer leaders, more selfish, and more inclined to tease (Item

27). (See Tables 5, 6, 7, 10, 11, 12, 13.) The former also used more hostility themes in their CAT stories than the latter.

The sex identification and intersib conflict we suspect may be greater the older the sister, within limits, and sufficiently strong to counterbalance other forces, such as we have enumerated above in the discussion concerning the boys, which should operate in the direction of reducing dependency and increasing friendliness to peers. The boy's chief models are females (the mother and sister); and the father, if Freudian theory is correct, may favor the sister.

A look at the probable characteristics of the older sister (see F1M's) shows she is likely to be relatively dominant and expansive, jealous, quarrelsome, exhibitionistic, insistent on her rights, critical, and competitive [see Tables 5, 6, 7, 8, 9, 10, 13; also see Items 18, 25, 31, 39, 42 (11)]. The qualities in the F1M's change little with spacing, a finding that may have some relation to the fact that the M2F groups at the various spacings differed little in attitude (Items 2, 4). If her brother were often left in her care, as well might be the case, the older sister could scarcely be expected to be a particularly wise and insightful guide. She would have a strong need to dominate and control. As evidence of stress between the sibs, we note that the M2F's [49-72]'s expressed a desire to change places with the sib, whereas, for instance, 87 per cent of the M2M [49-72]'s did (relevant to Item 27). The F1M's confessed with relatively high frequency that they thought they usually won in the quarrels with their brother. Forty-seven per cent of the M2F [49-72]'s thought the mother sided with the sib in the quarrels, whereas the corresponding per cent for the M2M [49-72]'s was 27.

We had expected that the older sister might be highly protective of her younger brother and, together with her parents, make the latter more adult centered and less skilled in interaction with children than would an older brother. The pattern of our ratings on the M2F's reactions to adults, as well as the interview data on sib interactions just mentioned, do not strongly support this hypothesis. For example, M2F [49-72]'s were not judged particularly friendly to adults (11), as one might have surmised. (This relatively low friendliness, of course, might be a reaction to domination, even though the individual is dependent and unprepared to take the buffets of his peers.) Nor do we find that a boy with a much older sister, who surely could be expected to be more maternal than a sister of nearly his age, has less positive peer attitudes than the boy with a slightly older sister. The M2F's, as was stated earlier, did not increase in sociability with peers as sib spacing increased, as was the case with the M2M's (Item 2).

The influence of a big sister taskmaster may be reflected in the high frequency of mentions of right and wrong—of moral issues—in the CAT stories of the M2F's. This frequency was consistently higher than was the case in the stories of the M2M's— in fact higher than in the stories of any of the other groups of second-borns.

A problem is posed by the fact that, while most of our male groups (exception M2F's) tended to become more friendly to and dynamic in relation with their peers as the maturity difference between them and their sib waxed, the female groups, especially the first-born, did not show the trend (Tables 5, 6, 7 and Items 4, 41). We suspect that the girl may have a limen of reaction to sib effect that is lower than the boy's. It may be less the character of the sib than the sheer fact of having a competitor for parent affection and attention that challenges. The girl is a conformist, jealous, and very anxious to please (21). Other ratings we have confirm this (11). Her character results, we think, in part because she identifies with her mother, the person who seems to her of most importance in the home and probably in her world (11). If this is the case, the sib competitor, but especially the one different in kind (sex), no matter what its age, is likely to cause her to feel challenged, to put her on her toes, to call forth a maximum response. The first-born girl, particularly the one with a brother, apparently defends her status jealously, as a king would defend his throne against a pretender, i.e., she is very active—more aggressive in fact than the usually more aggressive male—and this quality she shows, no matter what the spacing (see Items 13, 18, 25, 39, 42). We think she is as vigorous as she is because, in one sense, she has a better opinion of herself and feels more sure of her status than does the boy, this because of her identification with her mother. Stimulated and challenged she, among other things, reaches out to peers for attention and approval. Significant relative to the question of the greater peer orientation of the girl is the fact that girls in the interview reported having more playmates, especially when the sib was near in age.

As evidence for the young girl's tendency to have a better opinion of herself than the boy, we might mention Tuddenham's (21) finding. Working with young school-age children, this investigator noted that boys were inclined to rate girls higher on meritorious traits than they did boys and the girls did likewise. This state of affairs decreased with age. Also relevant to the issue is our observation that in the interview we had with our subjects it was the boy, particularly the first-born, who most frequently confessed he thought his mother favored the sib.

Apropos of the theory that boys, particularly those whose sib differs little in age, have less self-esteem and are more dependent than girls, let us mention several additional facts (relevant to Item 7). Our first-born male subjects at the closest spacing less frequently in the interview expressed a preference for male playmates than did the girl for female playmates and less often said they preferred to play with someone other than the sib than the girl did.

The second-born girl presents a somewhat different picture from the first-born, whose threat from the sib comes chiefly in the area of parent affection and approval. The girl with an older brother, as the sib-age difference widens, in contrast to the girl with a younger brother, seems to feel increasingly threatened *directly* along most lines (see Items 9, 11, 17, 31). Since the former has more opportunity to discover through the older brother that boys are given more freedom than girls and have a favored status culturally, and since the skill and knowledge difference between her and her sib is clearly in her brother's favor, we think she may have more sex identification and personal status conflict, as spacing increases—probably more than the girl with a younger brother. The girls with a much older brother were judged by their teachers as relatively tomboyish, for instance. Also at the middle spacing she expressed with the same group frequency as the F2F's (53 per cent) a desire to change places with her sib, although in her case, in contrast to the F2F's, the exchange of roles would involve a change in sex. It is significant, too, and probably reflective of her concern over her sib and herself in relation to him that F2M's CAT stories contained more sibling themes and mentions than those of any other group and the frequency of the sibling themes increased as the sib age difference widened (relevant to Items 9, 11, 15). In addition F2M mentioned with a higher frequency in her stories than did any of the second-borns, differences between persons—differences either with or without advantage to the person to whom they were attributed. This we think suggests a comparative set. F2M's bid more for adult attention at the wide spacing than close but also were judged less friendly to adults and less obedient (11). Still more significantly, our data reveal that the second-born girls with a brother increasingly as the sib age gap expanded, were thought to be less cheerful, more nervous, less responsible, more inclined to dawdle, more indirect in responses to fear and frustration, slower to recover from emotional upset, more vacillating, more likely to give alibis, and more apprehensive (data on preceding items not yet published). These seem to us to be signs that at the wider spacings stress was greater for the girl, or the response to it more depressive. The significant change seems to occur between the close and the two-to-four-year spacing, at which latter level the sib type (M1F) ratings also indicated for the sib the most change and this in the direction of greater aggressiveness, jealousy, hostility (Tables 5, 6, 8, 9, 11), but also social expansiveness (see Items 6, 16).

F2F's show a trend somewhat similar to that in the case of the M2F's, but not so marked, and apparently reach their peak of tension at the middle spacing (compare Items 9 and 10). But, of course, there are no readily recognizable signs of sex identification stress. The differences between closest and the widest spacings tended in the main to be small and insignificant. It is at the two-to-four-year age difference, when F2F is often cast in the role of a hanger-on, is dominated by her sister and her sisters' friends and has few friends of her own, that F2F shows most stress. At the middle spacing, in comparison with the close, F2F was rated less self-confident, popular (Tables 6, 7), kind, cooperative (Table 11), cheerful, planful, original, ambitious, and more selfish, vacillating, apprehensive generally, and given to teasing (Table 10), dawdling, and indirect responses to frustration (most data bearing on the items listed are not yet published).

It is interesting, when the two groups of second-born girls are compared, that those with a brother less than four years older tended to show the socially "positive" traits and competition to a higher degree than did the girls with an older sister (Items 15, 24 and Tables 5, 6, 7). This suggests that sex rivalry may be stimulating. At any rate, because of the sex distance factor (9), the F2M's at the middle spacing, we believe, have more friends of their own choosing than the F2F's—a situation that may also account for the greater effectiveness of the F2M's with their peer group. The latter will actually have less contact with the brother and his friends, because the breach between the sexes is greatest at ten to eleven years, than F2F will have with her older sister and the latter's friends.

At the middle spacing, when F2M is less of a hanger-on than F2F, the former can also be expected to be less hostile. Consistent with this extent-of-contact theory, we offer several other observations. At the widest spacing, for instance, when both groups of girls would have about equal amounts of sib contact, the sib age difference being sufficiently great to reduce the effect of sib's sex on the degree of association, the attitudes of our two groups of second-born girls were remarkably similar. Also girls with a brother their senior, with whom they probably associate relatively less than do first-born girls with their preschool brothers who would be at home most of the day, did not show hostile and competitive traits to so marked a degree as did the latter (Tables 8, 13). (See Item 31.) Of course it is not unlikely the second-born girl with a brother is relatively low in hostility, compared with the first-born girl with a brother, because the latter has her status as "first" to defend and has suffered the displacement experience. The generally higher rating of the F1M group in jealousy (Table 11), exhibitionism (Table 10), insistence on rights (Table 12), and sensitiveness supports this view (see Item 31).

Let us summarize here the forces which we think play on second-borns as sib spacing increases. The reader may wish to try to estimate their relative weights, for they do not all pull in the same direction.

I. Second-born boy with a brother:

a. This group doubtless has less sex identification conflict than M2F, and we would expect no great change with spacing in degree of conflict. If anything, a decrease might occur.

b. The mother's indulgence probably increases with spacing. The child in this group remains the youngest.

c. Increasingly with spacing the child's companions are of his age and choosing. His playmates are probably male increasingly, but possibly a peak is reached at the middle spacing. Only at widest spacing are many playmates his own.

d. The sib type of this group (M1M) is rated higher in social attitudes as the sib-age difference widens.

e. As the sib-age disparity expands, the child is outclassed increasingly in skill and wisdom by his sib.

f. The child gets increasingly more of his mother's attention uncontested the greater the sib-age difference, because the sib will be at school more of the time.

g. The mother is probably more relaxed the wider the sib spacing.

h. The child's identification with his sib is probably greatest at the close spacing (in the sense of feeling like the sib) but his desire to have the sib's assets and advantages may increase with spacing.

II. Second-born boy with a sister:

a. The sex identification conflict is doubtless strong in this group, since the mother and sib are opposite in sex to the child, and the conflict may increase with spacing at least up to the middle. The model the child has is strongly feminine, and it is not unlikely the father favors the sister.

b. e. f. g. are probably the same as in the case of I; but M2F may be more indulged, and his sib as well as mother may serve and protect him more. This may tend to unfit him for an effective role with his peers.

c. Increasingly with spacing the child's companions are of his age, sex, and choosing.

d. The sib type for this group (F1M) was rated very jealous and dominant. The sib type seemed to show few changes in attitude with spacing, except a low in responsiveness to adults at the middle spacing.

h. Probably a sharp reduction occurs with spacing in identification with, or *conscious* desire to be like, the sib.

III. Second-born girl with a brother:

a. The sex identification problem in the case of this group is great but probably not so great as in the case of II. The problem may increase with spacing. The mother probably favors the brother; the father, possibly the girl.

b. e. f. g. are probably as in I; but the child is not indulged so much as II.

c. Increasingly with spacing the child's companions are of his age and choosing. The child's playmates are probably female increasingly, but boys may be favored more than by F1M and F2F.

d. The sib type (M1F) was rated more sociable, less dependent, more dominant, aggressive, and hostile, as spacing increased. The peak of hostility seemed to occur at the middle spacing where F2M was also judged most quarrelsome.

h. The conscious desire to become the sib probably drops slowly.

IV. Second-born girl with a sister:

a. There is doubtless virtually no sex identification conflict in this group and little change, or possibly a slight decrease, with spacing.

b. e. f. g. are probably as in I; but we suspect the mother may favor the older sister.

c. Increasingly with spacing the child's companions are of her age and choosing, but the change is slow. F2F long remains a hanger-on in her older sister's play groups. Only at the widest spacing is F2F likely to have companions uniquely hers.

d. The attitudes of the sib type (F1F) are rated as changing little with spacing. The sib is likely to be distinctly jealous but less competitive than the sib in the case of III.

h. The childs' identification with the sib is very strong. The desire to have the sib's assets and advantages increases with spacing.

Since we are discussing in a general way some of the effects of difference in their ages upon the sibs, let us look more closely at the very great difference between first-born boys at the under-two and at the two-to-four-year spacing (see Item 6). The boy with a sister two to four years younger, as compared with one with a sister less than two years younger, was rated higher not only in friendliness, gregariousness, and social effectiveness generally, but also in hostility, jealousy, and competitiveness; while the boy with a brother his junior to the former degree, as compared with the latter, was rated higher chiefly in the hostile attitudes and techniques (see Items 6 and 24). The fact that the M1F [25-48]'s, compared with the M1F [7-24]'s, were both more friendly to peers and more jealous, quarrelsome, competitive, exhibitionistic, as well as more given to teasing, stimulates the conjecture that the socially "positive" qualities of M1F [25-48]'s are, in part, a defense. Whereas, for instance, 13 per cent of the M1M [25-48]'s expressed a desire to change places with the sib and 7 per cent of the M1F [7-24]'s did, 40 per cent of the M1F [25-48]'s confessed this inclination. This is impressive, since the exchange of roles in the M1F group would have to occur across sex lines, as well as represent age down-grading. However, we are impressed also with the drive or dominance of the M1F [25-48]'s. It should be remembered that with the sib much younger, the older child chooses their common playmates in large measure. These latter are more likely to be of the older child's age and sex than not, a fact which should bolster his self-confidence and may lie back of his active fight for attention and appreciation, for fight he must. The sib who is two to four years younger than a five- or six-year-old is at that adorable one-to-four-year age when it will probably be showered with attention and affection by adults, both in and out of the family. It is noteworthy that in all of the groups of first-borns the frequency of the wish to change places with the younger sib increased with spacing (data not yet published). The sib of one to four years gets about, in contrast to a baby who stays put, and may be very trying to the older child, interfering in a clumsy way with the latter's activities and constructs and offering not too much in the way of a compensating companionship. An infant sib would be less directly thwarting. It is interesting to note at this point that Stendler (20), with perhaps more emphasis on trauma and the "critical age" notion than we have presented, offers evidence which suggests that the arrival of a sib when the child is nine months to three years may be traumatic to the older one, interfering with the latter's dependency development. Sewall (19) also mentions the spacing of eighteen to forty-two months as that at which jealousy tends to be highest.

If, in addition, the sib of one to four years in age is a girl, she may seriously challenge M1F [25-48]'s sex identification, his basic self-evaluation. It is relevant to our hunch that a notable sex identification conflict occurs in this group, that, whereas 13 per cent of M1M [25-48]'s expressed indifference as to the sex of playmates, 53 per cent of the M1F [25-48]'s did and the corresponding percentages of preference for *male* playmates were 87 and 33, respectively. The boy, we suspect, feeling somewhat neglected and inferior, may try to lure attention to himself by a variety of devices, such as exhibitionism and teasing, and compensate for the devotion the sib receives by trying to build strong peer associations (Item 6).

Both M1M [25-48]'s and M1F [25-48]'s, let us repeat, were more hostile [quarrelsome, revengeful, cruel, given to teasing, and insistent on their rights (Tables 8, 12)] than corresponding groups at the close spacing; but the M1F [25-48]'s were, in addition, rated more friendly to peers, gregarious, popular, and higher in leadership (Tables 5, 6) as well as more jealous (Table 11), exhibitionistic (Table 10), and competitive (Table 13). (See also Items 6, 23, 24, 25, 26.) It seems clear that the middle spacing is a stressful one; but the M1F [25-48]'s apparently are dealing with their jealousy, perhaps partly because the sib could be excluded from their play groups, by an increase in sociability—by turning to peers—while M1M [25-48]'s are not doing this so much. From the peer groups which the M1F [25-48]'s join, since the sister is more than two years younger, she is likely to be excluded in large measure because of the cultural-biological phenomenon called sex distance (9). The boy's playmates will tend to be boys and they will help to reinforce his sex identification. If, in contrast, the boy's sib is a younger brother, he probably will not be barred from the older child's playmate groups, so peer associations can serve less as a compensation to the latter for the favors shown the sib by the family and others. The morale of the M1M [25-48]'s, in other words, probably will not be bolstered substantially by a coterie of unshared friends with whom he must stand on his own feet, and his defenses may be, hence, chiefly the more hostile ones.

Apropos of the above hunch, let us

Bureau Ednl. Psy. Research
[illegible] ING COLLEGE
Dated

call attention to the fact that M1M's and M1F's at the widest spacing do not differ significantly (see Item 24). When the spacing is wide enough—and over four years is adequate—then the great difference in maturity will bar the younger child from the older's play groups in large measure, regardless of his sex. Hence M1M's would have no advantage over M1F's and both groups could be expected to be, as they actually were, about equally friendly and gregarious. Also it is probably relevant that the sex of an *infant* as an item in the general social situation is less conspicuous than would be the case with a somewhat older sib.

It is interesting, in the light of our suspicion relative to the role of the sex identification conflict, that when the sibs differed by less than two years in age, males with a male sib who was younger were rated more friendly to children and adults than those with a female sib who was younger (Item 26). The male sib will not, because of his sex at least, challenge the first-born boy's status. He will be close enough in age, too, so that he will probably be a reasonably good companion, and identify some with his sibling. The two may even league against adults. These conditions should raise the self-esteem of boys with a younger brother near in age over that of boys with a closely spaced younger sister or that of boys with a brother two to four years younger. This constellation of forces should raise the independence of M1M [7-24], his self-confidence, his friendliness toward and interest in peers over that of M1F [7-24] (Item 26 and Table 5).

Below is a summary similar to the one given for second-borns of relevant influences which we think are playing upon first-borns.

V. First-born boy with a brother:

a. The child possibly has some sex identification conflict but, in the main, this is not very great. His conflict may be greater than M2M's because the latter has a male sib from the first to reinforce his identification. Possibly the closer the sib is in age the less conflict M1M has.

b. The longer the child remains the only child in the family probably the more indulged and possibly the more anxiously reared he will be.

c. The child is more adult-oriented the wider the spacing. He is likely to be more adult-oriented than M2M, especially at the closer spacings.

d. The child is more stimulated and instructed by adults the wider the spacing.

e. The boy's identification with his sib decreases slowly with spacing. The identification is not so great as M2M's, whose sib is older than he and commands more respect.

f. As the sib spacing widens, the child is ever abler than his sib.

g. As the sib-age difference expands, the child has less association with his sib, and the overlap in playmate groups diminishes slowly. At the widest spacing, when the sib is an infant or near infant, there is virtually no sharing of playmates. M1M's playmates at all spacings are chiefly of his choosing, in contrast to M2M's.

h. At the middle spacing the child perceives the sib as relatively more favored because the parents must interfere much in the children's affairs, yet the treatment differential must be great. The child may be less able to weather the storms of the displacement experience if it occurs at the "critical" nine-months to three-year age. We have the hunch the child may cling more if the sib arrives when he is one to two years old, because his defense repertoire is limited; and struggle more actively in defense of his rights and for attentions, if the sib's advent occurs when he is two to four years old. These attitudes we expect to transfer.

i. The child is probably given more responsibility than the second-born, M2M, at least when the spacing is close but is not given so much responsibility as VII or VIII because of cultural sex typing.

j. The sib type (M2M), according to our data, exhibits a better adjustment as spacing expands.

VI. First-born boy with a sister:

a. Much sex identification conflict tends to occur in this group. Possibly the peak is reached at the middle spacing where the treatment differential may seem most conspicuous. The sex of an infant sib, for instance, is not very conspicuous and may not be perceived as responsible for differences in treatment. The mother may favor the boy; the father, the sister.

b. c. d. e. f. g. h. i. are probably much as in V. However, the identification with the sib and also the amount of association with the sib and her friends doubtless decreases more rapidly with spacing than is the case in V. Also the child is probably given more responsibility than is M2F but at the close spacing at least not much more than his younger sister.

j. The sib type (F2M), according to our findings, is actively defensive when the spacing is under four years but seems relatively more passive and dependent at the widest spacing.

VII. First-born girl with a brother:

a. Much sex identification conflict seems to be experienced by this group; but it *may* be less than that in VI because of the child's identification with the mother. However, increasing relative interest in the father may intensify the girl's conflict. Also her displacement experience is likely to involve a more radical change in status for the girl than for the boy in V or VI, as the younger brother tends, we think, to be much indulged by the mother. This state of affairs should put F1M much on the defensive at all spacings.

b. c. d. f. seem much the same as in V. However, the girl, because of her identification with her mother and teacher, is more adult-oriented than is the boy.

e. Identification with the sib probably decreases more rapidly with spacing than in VI or VIII.

g. As the sib-age difference expands, the child has less association with her sib, and the overlap in playmate groups diminishes but probably is greater than in VI.

h. Since the advent of the sib makes for a greater change in status for her than for V or VI, for instance, her defensiveness may vary less with spacing and be very active.

i. The child is given more responsibility than F2M and much more responsibility relative to the sib than V or VI. Her responsibility for the care of the sib is likely to increase with spacing. Since, as a girl, she is much of a conformist, she may make much effort to keep her sib on the straight and narrow path and be much upset if he resists. Hence sib relations may not be too good.

j. The sib type (M2F) tends to be relatively dependent and hostile and changes little with spacing.

VIII. First-born girl with a sister:

a. The sex identification conflict in

this group is probably inconsiderable. It may be slightly greater the wider the spacing.

b. c. d. f. are as in VII.

e. Identification with the sib tends to remain high but lessens slowly with spacing.

g. As the sib-age difference expands, the child has less association with her sib, but the overlap in playmate groups remains high until at least the widest spacing.

h. The advent of the sib will not change her status as much as it does that of VII and the sib treatment differential will remain less because the sibs are of the same sex. It is not unlikely F1F will remain her mother's favorite.

i. The child is given more responsibility than F2F and may have less difficulty keeping her sib in line than VII does. The middle spacing seems most difficult.

j. The sib type (F2F) seems to change little with spacing, though the signs of stress pile up at the two-to-four-year spacing.

Before turning attention to variables other than spacing, we wish to comment briefly on the several traits related rather simply to spacing—i.e., leadership (Table 6), competitiveness (Table 13), and teasing (Table 10). (See Item 5.) That leadership is positively correlated with the degree of difference in age on the part of the sibs may reflect conditions already mentioned—namely, that the child whose brother or sister differs from him much in age tends to be thrown more with children of his own age, sex, and choosing and hence have more motivation and opportunities to lead. Also if the sib is junior, the older child may be given relatively more responsibility the younger the former is; and because of his obvious superiority to his sib in skill and knowledge, get practice in leading.

That teasing should increase as sib age disparity widens, at least from the under-two-year to the two-to-four-year spacing level (Item 5), seems to us to be expected, it being a favorite method of attack to use when more direct forms are forbidden or unsafe. The child is likely to be forbidden to attack a younger sib directly and this to a greater degree the younger the sib is. Hence teasing should be on the up-grade, except that an infant would be a poor target for teasing, and would also be protected against even that "playful" form of attack. This latter we think lies back of the fact teasing in first-borns dropped to some extent at the widest spacing, when the sib in the population we studied would be virtually an infant and might not even understand it was being attacked. If the sib, on the other hand, is older than the child, it will be unsafe to attack him directly. Since the second-born child is outclassed physically by his senior sib, the former will be inclined to use those "safer" kinds of attack of which teasing is one. Presumably because boys are more aggressive, they use teasing more than girls (Item 38) and also against girls, with whom they are taught it is unmanly to fight directly.

Competitiveness seemed to increase some as the age difference between the sibs expanded, at least up to four years (Table 13). This surprised us at first, as we had expected the trait to be greatest at the closest spacing where overlapping interests and frequency of association are greatest. We suspect now that status and affectional concerns, rather than, for instance, concern over possessions or skill, are the chief sources for the motivation of competitiveness. Probably the longer the first-born has held the family stage uncontested the greater will be his need, if challenged, to prove himself best. An infant sib of not very long standing, however, probably will not challenge much. The sib at the middle spacing is more threatening.

Among second-borns it was those with male sibs that tended to increase in competitiveness with spacing. The apparent great increase in hostility, expansiveness, and competitiveness of these very male sibs (see sib type M1M and M1F and Item 6) probably gives us the answer to the increase in competitiveness in the second-borns, whose sibs the latter are.

Sib's sex. That our first-borns tended to be more rivalrous and hostile if the sib was opposite in sex (Items 24, 25), whereas the reverse, if anything, obtained for the second-borns, presents a problem. This suggests to us the sex identification struggle and degree of sib association are two significant determinants, the former variable being relatively more significant for first-borns; the latter, for second-borns (see Items 31, 33). If the sib is himself in the Oedipal stage or not well established in his sex identification, then he activates insecurity in the five- and six-year-olds, who have probably not yet entirely resolved their own sex identification problem. Also, if one of the sibs is a preschooler and the other has just entered school, then the two will have been thrown in each other's company much and have played with the same playmates, unless one of the two is an infant. There may be much sex rivalry in mixed groups of playmates. If, however, the sib is older, and well established in his identification, then sib rivalry will probably revolve relatively more around other issues. If the sib is older, sex distance will be great. This means that the younger child will have considerably more contact with his older sib and the latter's associates if the latter is of his own sex rather than opposite in sex. Hence we should expect reflections of this in relatively more rivalrous attitudes toward sib and peers among members of same-sex pairs (see Item 33).

The rivalry of girls with a younger brother, as compared with girls with a younger sister, another relationship to be observed in our data (Items 23, 25), is conspicuous. The former girl because of sex rivalry, we think, is highly competitive and is apparently sufficiently self-assured (she is rated high in self-confidence by her teachers, for instance) to express her hostility directly. These girls with a younger brother were at the closer spacings, as has been already described, even more quarrelsome than boys with a younger brother (Table 8, Item 42). The former were judged better leaders (Table 6), more quarrelsome (Table 8), competitive (Table 13), exhibitionistic (Table 10), jealous (Table 11), insistent on rights (Table 12), and critical (Table 9) than girls with a sister. That the boy, who has this self-assured, very jealous, dominant older sister (see obverse group, F1M) whose attitudes seem little affected by the age of her sib, should also be little influenced by spacing may be significant. The boy is not very expansive socially but instead rather hostile. He, doubtless, feels inferior and dominated, as has already been explained, and withdraws. He may, however, be compensated for his, in one sense, increasingly inferior position by enhancement in the babying and protection he receives. While he ranks consistently lowest of all of our groups in friendliness to children, it is significant that at the middle spacing at least he ranked highest in friendliness to adults (11).

One other of the sib's-sex relations let us single out for comment, though it has already been indirectly referred to. In the main, children with a brother (exception first-born boys) tended to be rated higher in competitiveness and leadership in comparison with those with a sister (Item 23).

Since the boy was not judged a better leader or more competitive generally than the girl, sib example can scarcely be the explanation. Sex rivalry, we think, accounts chiefly for the difference between the FM and FF groups. However, that M2M's exceeded M2F's may have a different explanation, as was suggested earlier, though the mechanism may be related in part to the sex identification conflict also. The boy with an older sister we believe is overwhelmed by his double disadvantage of being younger and not of his mother's and sib's sex; and, depressed by this and the excessive protection he receives, is dependent rather than stimulated. The boy with an older brother, on the other hand, who has his sex identification reinforced by the sib and the latter's friends, can, at least at the wider spacings, be expected to be more dynamic and buoyant (Item 7).

Sex. Since the sex variable has been dealt with fairly adequately in our discussion of its interaction with other variables, comment here can be rather brief. Sex differences do not seem simply related to the variables studied—a fact which suggests that the biological sex vector can be countered in influence by cultural or sex-role influences. Only in uncooperativeness (Tables 11, 20) and tendency to tease (Tables 10, 19) did boys with any consistency exceed girls in the traits studied (Item 38). Even here at the close spacing there were no significant sex differences, and among second-borns, sex differences were significant only between those children with a sister. Among first-borns, boys were judged more uncooperative than girls when the sibs came from opposite-sex pairs. Neither among first- nor second-borns were any significantly large sex differences shown if the children belonged to sib pairs of the same sex. When children from opposite-sex pairs are compared, boys tended to exceed girls in uncooperativeness. A similar pattern is to be observed in the case of teasing (Tables 10, 19), though when the sib-sex variable is kept constant, boys rather uniformly averaged higher than girls (Item 38).

Girls were more self-confident, more friendly, better sports, possessed of more initiative, and were socially more effective than boys only at the close spacing (Tables 5, 6, 13; Item 40). At this age-difference interval the girl's impact with the sib and his friends is most direct. Our hunches relative to the meaning of the general observation given above we have described elsewhere in terms of the greater likely indulgence of the boy by his mother and his consequent greater dependence; the girl's greater self-assurance because of identification with her mother; and her tendency to deal with struggle more indirectly—a condition which might cause her to reach out to peers, if the sib is not too satisfying, and thus develop more social experience. However, this does not mean the first-born girl is not competitive. She is sufficiently challenged and stimulated by rivalry across sex lines with her younger brother; she was rated for instance more quarrelsome (Table 8), competitive, and critical (Tables 13, 9) than the parallel boy group (Item 42). That it is sex rivalry that is operating is suggested by the fact that the girl with a younger brother is more competitive (Table 13), quarrelsome (Table 8), jealous (Table 11), exhibitionistic (Table 10), insistent on rights (Table 12), and critical (Table 9) than one with a younger sister (Item 42). The former's involvement with her brother is suggested by the greater frequency of sibling mentions in her CAT stories.

Ordinal position. We were impressed with the fact that few of our dependent variables were related in any simple way to ordinal position. This finding suggests that the sweeping generalizations found in the literature about the effect of birth-order may need some tempering.

In the main, our first-borns at the wider spacings were judged, in comparison with second-borns, more competitive (Item 30), insistent on their rights and curious. In their CAT stories the former used more hostility and favoritism themes (data not yet published). These may be a reflection of the first-born's need to defend his status as "first," which in turn may hark back in part to the displacement experience. In keeping with this hunch is the observation that, at least at the stressful middle spacing, the first-born is the more upset by defeat (Item 32). It is interesting, however, that at the close spacing our second-borns when the sib was of the same sex were more insistent on their rights than the parallel group of first-borns, while the reverse obtained for those from opposite-sex pairs (Item 33). And boys with an older brother were rated more quarrelsome and hostile generally than boys with a slightly younger brother (Item 34), whereas for girls with a brother the opposite seemed to be the trend (Items 31 and 42). Having a competitor from the very first and doubtless also being encouraged to defend himself if the sib is not much older, probably makes M2M a better fighter than M1M. The girl with an older brother, overpowered by him and probably not encouraged particularly to defend herself directly may resort more to indirections than F1M and hence appear less hostile. At the wider spacings M2M, like F2M at the somewhat closer, we would expect to be convinced of the hazards of direct attack and hence to be more cautious and indirect. Our data show, for instance, that, in contrast to M2M (7-24), M2M (49-72) tended to bid for more help and attention from adults (11).

It seems to us particularly provoking of thought that ordinal-position group differences were few at the close spacing and when the sibs were of the same sex, but frequent when the sibs were opposite in sex. A similar trend appears in the frequency of sibling mentions in the CAT (2) stories. We have, of course, already looked at the reverse side of this picture in commenting that heightened social interest, friendliness, insistence on rights, competitiveness, jealousy, exhibitionism, and quarrelsomeness seem to obtain among first-borns in opposite-sex sibling pairs more than among second-borns, whereas among the latter the reverse, if anything, seems more characteristic (Tables 8-13). The fact that there is a difference in favor of those from opposite-sex pairs among the former makes us suspect jealousy and unresolved sex identification problems are sparking the drive of the first-borns. This is no doubt reinforced by the knowledge on the part of these latter children they are the abler members of the sib pairs. We have thought it important, too, in this connection, that among second-borns, sibs of the same sex associate with each other more than do those of opposite sex; whereas among first-borns the two sexes tend to play much together, unless the sib is very young. Our interview material supports this surmise. The playmates are, to be sure, more of the older child's choosing, since the younger child, being usually under four years of age, is too young to get about much on its own and to contact playmates. The strong peer interests of first-borns we think may stem in part from the inability of the younger sib, unless near in age, to be much of a real companion. It is significant that in the interview we had with the children first-borns somewhat more frequently than second-borns expressed a dislike of playing alone but also with higher frequency said they preferred playing with someone other than the sib. The first-born apparently reaches out for stimulation, especially when the sib is much younger and can give him little real intellectual companionship.

The sex identification of the first-born is likely to be reinforced by his playmates but his problem is probably still not completely resolved because of the amount of attention the sib gets from adults. Since, if the sib is younger, neither the child nor the sib has fully accepted its sex and sex-role implications, rivalry may be great if the sibs are opposite in sex. The first-born, it is our belief, gives a relatively dynamic, direct response to the situation because he is in the driver's seat. In other words, he is self-confident enough to be directly hostile. He may be more rivalrous than the second-born also because of his displacement experience and the feeling he has his status as "first" and ablest to defend. The M1F's, too, may be subject to a devastating sort of comparative criticism, especially if the sib is near in age. The sister may well be a more conforming sort of person (11)—little girls tend to be—and this may reflect more unfavorably on the boy than would be the case if the younger sib were a relatively less conforming male, for instance.

Congruent with the above hypothesis is the observation that M1F [7-24]'s were rated as less friendly to adults, less affectionate (11), less exhibitionistic (Table 10) and less inclined to seek adult attention than M2F [7-24]'s, while M1M [7-24]'s did not differ similarly from M2M [7-24]'s in these traits. When the situation with the sib is not quite so taxing for the M1F's, as is the case when the latter become clearly more able than the sib and can have play groups all their own, they may compensate for what inferiority feelings any comparative criticism and their sex status generate by effort to win peer approval and attention. M1F's after the closest spacing were judged more friendly to peers than M2F's (Item 31) and about as friendly to adults (11). That this is no sign of peace is suggested by the observation M1F [25-48]'s were gauged more quarrelsome, revengeful, competitive, insistent on their rights, given to teasing, cruel, and jealous than M2F [25-48]'s. F1M [25-48]'s were judged more insistent on rights, competitive, reactive to defeat than F2M [25-48]'s, while F1M [49-72]'s were rated more on the defensive—i.e., quarrelsome, exhibitionistic, attentive to adults, and demanding of help than F2M [49-72]'s. (See Tables 8-13; also Reference 11.)

Worthy of some comment is the greater hostility (quarrelsomeness, insistence on rights, selfishness) of M2M [7-24]'s than M1M [7-24]'s (Tables 8, 9, 12; also Item 34). The boy with a brother slightly his senior is doubtless subjected to much domination by the latter with whom he plays much. However, the M2M [7-24]'s are near enough to the M1M [7-24]'s in ability to hold their own if they make a strenuous effort. Since the former will not get the protection a much younger child would receive and will be encouraged by parents to defend themselves, their more directly combative habits will probably be those most reinforced. At the middle spacing then, when the M2M's will be given more protection against the older sib, it should be the M1M's who are most hostile and on the defensive.

That at the middle spacing the M2F's did *not* exceed F2F's in hostile attitudes [quarrelsomeness, cruelty, revengefulness, exhibitionism (Tables 8, 9, 10; Item 43)] we think may be due in part to a difference in degree of association between the sibs. At the close and wide spacing, when the differential in the degree of association between the members of mixed as opposed to same-sex pairs is small, the M2F's exceeded the F2F's; but at the middle spacing, when the members of the M2F pairs associate less than those of the F2F pairs, there is an influence which counters the hostility strength vector of the boys.

The patterns described above suggest again that having a sib opposite in sex tends to be stressful. The boy doubtless uses teasing and resistance to defend himself and his status. Since he will be deterred from attacking the girl directly more than the girl will be deterred from attacking him, we should expect to find what appears to be the case; namely it is when the sib is a sister that the sex differences in teasing are most conspicuous (Items 12, 43).

Girls generally had higher mean ratings in jealousy than did boys (Tables 11, 20), sex differences being largest at the widest spacing (Item 39). Girls with a younger brother tended to be rated highest in jealousy. Also in their CAT stories they showed a more comparative set. The general direction of the sex difference we noted is in line with that reported by others (e.g., see 7, 15, 19). The girl is more of a conformist, for whatever reason, than is the boy, more affectionate, and probably also the recipient of more affectionate expressions (11). She wants approval and when she sees others getting it, she is motivated to try for her share or, failing this, to justify to herself her apparent failure to capture it. The girl seems more involved with her family. In her CAT stories there were more mother, father, and sib mentions than in the boys' stories.

That the sex rivalry problem is a factor influencing her jealousy is suggested by the fact that among first-borns the sex differences in the trait are significant if children from mixed-sex sib pairs are compared but not if those from same-sex pairs are compared. This is not the case among second-borns, where, as we pointed out earlier, those of the same sex associate more and hence probably are more rivalrous. Because the girl is more of a conformist, a sib who differs from her in so basic a character as sex may cause her to challenge her worth more than a sib of the same sex. It is possible, though it seems unlikely, that the girl even in her preschool years has learned she is living in a society in which the male is favored.

Obverse groups. We had hoped when we began the study to concern ourselves with both members of each sib pair but this did not prove feasible. Hence we have had to trust the picture in the obverse group to give us some idea of what the trend among the actual sibs of the children in any given subgroup would be. There is even a compensating feature in the obverse group technique in that the members of sib type groups are the same in age as the members of the comparative subject group. Since we believe the age variable an important determinant of the trait pattern to be observed among sibs from various constellations, let us warn the reader once again of the danger of generalizing from our findings to other age levels.

We have looked at the obverse groups in two ways—one, in terms of their relative ranking in total score. When we observe, for instance, that F1M is very jealous and M2F is low (Item 47) we suspect F1M may have a reason to be jealous, namely that M2F is babied and protected much. However, this negative relation may reflect merely the influence of a sex variable, quite apart from the sib influence. Such a variable may cause the difference between the sexes in opposite sex pairs. Hence we have looked at another set of relations for cues as to the role of direct sib influence—namely in the pattern of change with various variables—e.g. spacing. There is a suggestion—but of this we are not very sure—that if a subject type does not change in essential character with spacing the obverse group tends also not to change significantly. First-born girls, for instance, changed little in friendliness to peers with spacing. The same pattern obtained in the obverse second-born groups, F2F and M2F.

IV. SUMMARIZING COMMENT

Since any adequate summary of the findings of the study would repeat what is given in Section II, the reader is referred there for a review. Because the detail of our data is complicated but, in the main, has made sense, we have been impressed with the dangers of broad generalizations concerning the effect of family constellation variables. Our most significant findings are probably to be located in the various interactions. We suspect, too, that the pattern of some of our findings is a function of the age group we studied, and wish to remind the reader once again of this possibility. We are inclined to point out as highlights in our material, evidence suggesting (*a*) that the two-to-four-year spacing may be a rather stressful one; (*b*) that having a sib opposite in sex may be very stimulating and security taxing, especially to first-borns; (*c*) that a boy with a much older sister tends to be somewhat withdrawn and dependent (though still distinctly hostile) in comparison with a boy with a much older brother; (*d*) that the social expansiveness of most male groups tends to be positively correlated with degree of disparity in age (within the limits we studied) of the child and his sib; (*e*) that differences associated with ordinal position are frequently contingent on sex of child and sex of sib; and (*f*) that the variable, sex of child, interacts extensively with the sibling variables investigated. Simple or "zero order" sex differences, for instance, were noted in our data extremely infrequently.

REFERENCES

1. ADLER, A. Characteristics of the first, second, and third child. *Children,* 1928, **3**, 14, 52.
2. BELLAK, L. & BELLAK, SONYA. *Children's Apperception Test.* (2nd Ed.) New York: C. P. S. Co., 1950.
3. BERNE, ESTHER V. C. An experimental investigation of social behavior patterns in young children. *Univer. Iowa Stud. Child Welf.,* 1930, **4**, No. 3.
4. CONRAD, H. S. *The California Behavior Inventory for Nursery School Children.* Berkeley: Univer. of California Press, 1933.
5. CONRAD, H. S. A statistical study of ratings on the California Behavior Inventory for Nursery-School Children. *Genet. Psychol. Monogr.,* 1934, **16**, 1-79.
6. EDWARDS, A. L. *Experimental design in psychological research.* New York: Rinehart, 1950.
7. FOSTER, S. A study of the personality make-up and social setting of fifty jealous children. *Ment. Hyg.,* 1927, **11**, 53-77.
8. GOODENOUGH, FLORENCE L., & ANDERSON, J. E. *Experimental child study.* New York: Century, 1931. Pp. 501-512.
9. KOCH, HELEN L. A study of some factors conditioning the social distance between the sexes. *J. soc. Psychol.,* 1944, **20**, 79-107.
10. KOCH, HELEN L. The relation of "primary mental abilities" in five- and six-year-olds to sex of child and characteristics of his sibling. *Child Develpm.,* 1954, **25**, 209-223.
11. KOCH, HELEN L. The relation of certain family constellation characteristics and the attitudes of children toward adults. *Child Develpm.,* 1955, **26**, 13-40.
12. LASKO, JOAN K. Parent behavior toward first and second children. *Genet. Psychol. Monogr.,* 1954, **49**, 97-137.
13. MAYER, A. *Chicago: Socio-economic groups of the population by census tracts.* The Chicago Community Inventory. Chicago: Univer. of Chicago, 1940.
14. RICHARDS, T. W., & POWELL, MARJORIE. The Fels Child Behavior Scales. *Genet. Psychol. Monogr.,* 1941, **24**, 259-311.
15. ROSS, B. M. Some traits associated with sibling jealousy in problem children. *Smith Coll. Stud. Soc. Work,* 1931, **1**, 364-376.
16. SANFORD, R. N., *et al.* Physique, personality and scholarship. *Monogr. Soc. Res. Child Developm.,* 1943, **8**, No. 1.
17. SEARS, PAULINE S. Doll play aggression in normal young children. Influence of sex, age, sibling status, father's absence. *Psychol. Monogr.,* 1951, **65**, No. 6 (Whole No. 323).
18. SEARS, R. R., WHITING, J. W. M., NOWLIS, V., & SEARS, PAULINE S. Some child rearing antecedents of aggression and dependency in young children. *Genet. Psychol. Monogr.,* 1953, **47**, 135-236.
19. SEWALL, MABEL. Two studies in sibling rivalry. I. Some causes of jealousy in young children. *Smith Coll. Stud. Soc. Work,* 1930, **1**, 6-22.
20. STENDLER, CELIA B. Possible causes of overdependency in young children. *Child Develpm.,* 1954, **25**, 125-147.
21. TUDDENHAM, R. D. Studies in reputation: I. Sex and grade differences in school children's evaluations of their peers. *Psychol. Monogr.,* 1952, **66**, No. 1 (Whole No. 333).
22. WHITING, J. W. M., & CHILD, I. L. *Child training and personality.* New Haven: Yale Univer. Press, 1953.

(Accepted for publication December 8, 1955)

No. 427 -28
1956

Vol. 70
No. 20 -21

Behavioral Differences Between Normal and Traumatized Newborns:

I. The Test Procedures

By

Frances K. Graham
Washington University

Bureau Ednl. Psy. Research
DAVID HA ING COLLEGE
Dated ....
Acc. No ....

Vol. 70, No. 20 Whole No. 427, 1956

Psychological Monographs: General and Applied

Behavioral Differences Between Normal and Traumatized Newborns:

I. The Test Procedures[1]

FRANCES K. GRAHAM
School of Medicine, Washington University

THE present work was undertaken as a necessary steppingstone in moving toward the solution of a more complex problem. Many medical disciplines have been concerned with "birth injuries"[2]—the factors giving rise to them and their consequences, both immediate and delayed. In studying the causes of birth injury, it has become clear that they include much more than the conditions under obstetrical control, as the anesthesia, type of delivery, etc. The status of the infant himself, his size, maturity, and resistance, are of at least equal importance. Similarly, in attempting to assess the consequences of birth injury, one immediately finds grounds for the suspicion that a given external trauma will have different effects depending on such factors as whether the newborn is full term or premature, what the placental placement was, the mother's state of nutrition, etc. The number of probably significant variables that have been implicated is so large that it has seemed as if only a large-scale statistical study involving thousands of cases could possibly cope with the complexity.

The author felt that another method of attack would be necessary to answer the highly important practical question of *which* infant will show later abnormalities in development. This method was to begin with the infant rather than with external measures of trauma. If factors inherent in the individual infant introduce so much of the cumbersome complexity, then ordering infants according to the severity of their response to trauma might better predict the outcome. This is, in fact, the procedure which is employed by the physician in his clinical capacity. He makes clinical judgments or predictions about an individual infant, based on knowledge of the trauma, but based primarily upon how the infant himself responds after birth. In order to make this process scientifically useful, it is necessary to find some way of quantifying or ordering the clinical judgments and, preferably, of making them independent of the skill of particular practitioners.

---

[1] This work was initiated by a four-month grant from the Frances Israel Fund of the Noshin Rachmonioth Society of St. Louis, Mo. From Nov. 1, 1953 to Oct. 31, 1954 it was supported by the Children's Research Foundation, St. Louis, Mo., and since Nov. 1, 1954 by research grant B685 from the National Institute of Neurological Diseases and Blindness, of the National Institutes of Health, Public Health Service. It is part of a long-term project to study the effects of anoxia, being carried out by Drs. Alexis F. Hartmann, Miriam M. Pennoyer, and the author, as principal investigators. Appreciation is expressed to the many others who have participated in the project, and especially to Drs. Ruth G. Matarazzo and Bettye M. Caldwell who have contributed to the psychological aspects of the research.

[2] Birth injury is defined broadly to include any condition at birth giving rise to brain injury, whether primarily from anoxia or from mechanical trauma.

Most important from the point of view of the consequences of birth injury, is the desirability of distinguishing the infant whose response to asphyxia is severe from one whose response is mild or absent. It has been argued on the one hand that an infant who survives the experience of anoxia at all will survive without later consequences and, on the other hand, that a large number of psychiatric and intellectual difficulties in later life may be ascribed to this cause. The literature is extensive as well as controversial but a brief summary of major studies may be found in Ford's (5, p. 951) authoritative text. It would seem that more agreement as to eventual consequences might be reached if the neonatal response were taken into account. One would expect that the later sequelae would not be the same for the infant whose condition was precarious for several days after birth as for one who appeared normal even though subjected to apparently equal trauma.

Another advantage of approaching this field by studying the infant's response is that it is possible to obtain information immediately concerning a number of problems, rather than waiting for the passage of years. Just as the mortality rates with various obstetrical procedures are of significance, so is it of interest to know how or whether the condition of a newborn infant asphyxiated for five minutes differs from one who has not been so traumatized. It is of interest to know whether use of obstetrical methods such as natural childbirth are reflected in any differences in neonatal behavior. Where no immediate differences can be found, it may be fruitless to search for them at a later stage when the number of other possibly relevant variables has increased a hundredfold.

While the desirability of considering the infant's response to trauma may be readily acknowledged, it is necessary to develop a satisfactory method of measuring newborn behavior before this variable can be introduced into experimental work. The present paper reports our efforts to do this. We have spoken of measuring "response" or "behavior," but this is an amorphous categorization which may be variously defined. For our present purpose, we are interested in measuring any response in the behavioral repertoire of the neonate which may be related to the kinds of trauma and the kinds of consequence which may be the result of birth injury. More specifically, we set as our goal the measurement of behavior which would differentiate a group of infants who were "normal" from those who might be candidates for "brain injury." We were interested primarily in differences between normal and abnormal or traumatized infants and not in variability among normal infants.

There have been a number of systematic, psychological studies of newborn behavior but these have been restricted to normal infants. A review of these studies may be found in Pratt (13). General descriptions of the behavior of "abnormal" newborns are available in pediatric textbooks, and Gesell and Armatruda (6, p. 215 ff.) have given detailed reports of individual brain-damaged cases. This literature provided clues as to the kinds of responses which might be usefully measured, and we attempted in preliminary work to develop objective measurement techniques. Five procedures were eventually developed—measures of pain threshold, maturational level, visual response, and ratings of irritability and muscle tension. It is the purpose of this and the following paper to report in detail the methods and their standarization and degree of differentiating power. These five tests constitute a measure of neonatal response now being emphasized in a second study as a basis for prediction of subsequent development.

TEST PROCEDURES

In selecting these particular procedures, we were guided by a number of considerations. The primary one was that the test should differentiate normal from traumatized *S*s and that the range of possible variation should be sufficiently great so that differences would not be minimized by a low ceiling. A second consideration was the desirability of making measurement objective and reliable. We did, however, prefer to explore relatively subjective measures giving promise of validity rather than concentrate attention on developing precision of measurement in an area where the worth of measuring at all remained to be established.

It was also necessary to take into account special difficulties arising from the nature of the *S*s and the hospital setting. In the infant, for example, the momentary internal state is a factor which may obscure the effect of other determinants of behavior if allowance is not made for it. In general, we attempted to solve this by maintaining a constant state between waking and sleeping. The specific methods of dealing with the problem are described under the several procedures.

The infant also presents special difficulty in that his repertoire of response is limited and that there seem to be few clear-cut stimulus-response connections. A great deal has already been said about generalized mass movements as characteristic of the newborn. Whether he characteristically indulges in generalized or in specialized movements, it is unquestionably difficult in many cases to determine whether movements occurring after application of a stimulus are "responses" or are spontaneous. One of our major tasks, therefore, was to discover responses which did show a specific relationship to a stimulus and to set up rules for unequivocally identifying the response.

Changes in behavior during the newborn period are very rapid, and behavior characteristic of Day 1 may not be present to the same degree on Day 2. Either the age of infants must be held constant to the day, or the relationship between age and the particular test measure must be empirically determined. In order to obtain maximum differences we wanted to examine infants on the first day. Yet the most severely traumatized *S*s, the most likely to show marked differences, often could not be tested for several days. Holding the day of testing constant, therefore, meant either sacrificing the sensitivity gained from testing on Day 1 or sacrificing the *S*s most severely traumatized. Rather than make either of these sacrifices, we carried out the more time-consuming method of obtaining age norms.

Other considerations more related to the nature of the project than to the nature of the *S*s had to be taken into account. The fact that we were dealing with hospital patients, many of whom were seriously ill, necessitated the use of procedures which would be minimally disturbing to the *S*. Considerations of space, subsequent clinical use of tests, and the scope of the project disposed us to seek tests which did not take long to administer and did not require elaborate apparatus.

The most seriously limiting factor was the relatively few abnormal *S*s available. In the well-run, modern obstetrical hospital, the percentage of babies who suffer severe birth injury is only 1 to 3 per cent. This meant that an immediate check on the validity of a measure could not be made but had to wait until a sufficiently large number of traumatized *S*s had accumulated. Probably the major effect of this was only to threaten the experimenters' morale, but it also meant that once a test procedure had been adopted and given to any appreciable number of abnormals, changes which might have improved the procedure were not made since they entailed too great a loss of difficult-to-replace data. It was because of this scarcity of traumatized *S*s that we also included infants with mild or questionable injury rather than using more widely separated criterion groups. Except for the apparatus used in the measure of pain threshold, all materials used in this study are readily available and could easily be duplicated by others conducting research in the same area. A detailed description of the exact procedure for each technique is given in the following pages in order to facilitate such usage and to provide the necessary information for interpreting subsequent results.

Pain Threshold

A determination of pain threshold was employed for a number of reasons. Both pediatric and psychological examinations of the newborn have concentrated on motor ability and have neglected sensory functioning. Yet impairment of sensory functioning has been a useful sign of brain damage in the adult, and Windle (17, p. 57) reports that, after intrauterine asphyxiation of guinea pigs, sensory defects, when determinable, were more persistent than motor defects. We selected pain as the sensory area on which to concentrate precise measurement, after preliminary experimentation revealed various practical difficulties in the other areas. One major advantage of using pain was that a specific, discrete response could be obtained which was relatively easy to differentiate from spontaneous movement. Another consideration was the possibility that sensitivity to pain has not reached a maximum at birth. If so, pain thresholds might prove a relatively sensitive indicator of brain damage, since many lines of investigation have suggested that functions which are not well established are especially susceptible to impairment. Previous studies of sensitivity to pain in newborns have recorded any response that could be observed rather that a specific response. They have also varied the method of stimulation and have not used stimuli of graded intensity.

Our general procedure was to determine the stimulus intensity necessary to elicit a specified response within a specified time. Initially an ingenious method of mechanical stimulation developed by Bishop (2) was tried. Despite the superiority of this apparatus over previous mechanical methods, control of the rate and angle of application was found to be too difficult, at least in working with infants, and the procedure was abandoned. An electronic stimulator, also developed by Bishop (3), was substituted and found satisfactory. The apparatus delivers shocks whose intensity, duration, and frequency can be varied within the stimulating range for skin and peripheral nerves. For the present work, duration and frequency were held constant so that the shock consisted of a two-second impulse at a frequency of 14 per second which could be varied in intensity from 50 to 530 volts. Eleven voltages covering this range and corresponding to the points on a control dial were selected as stimuli. The average difference between stimuli was 50 volts. A resistance of one-half meghom was in series in the circuit, thus reducing the effect of individual skin resistance. We also made separate measurements of skin resistance in a number of *S*s to check that, in fact, the resistance did not determine the response.

The infant was stimulated while lying on his back with the legs in the usual flexed position. The indifferent electrode, covered with moist gauze, was placed in the center of the back against the skin, and the shirt tied so that it remained in position. A circular area, approximately ¾ inches in diameter, just below the knees of both legs was treated with vaseline. These areas were stimulated alternately by a silver ball electrode of 1 mm. diameter. An effort was made to avoid stimulating the same spot successively, especially if the skin reddened. There was apparently sufficient spread of stimulation so that differences in spot sensitivity did not introduce undue variable error.

The specific response required was a movement of the stimulated leg or the foot of that leg, with or without other bodily movement. Wickens (16) has described a similar specific response when shock was applied to the soles of the feet. While the characteristic response to the stimulus was a raising of the stimulated leg only, with foot flexion and either flexion or extension of the knee, there was considerable individual variation. Our only requirement, was, as noted above, that the stimulated limb should move. If there was any doubt as to whether or not the movement was spontaneous, as in cases with unusually short latency or where the movement appeared to be part of a generalized movement such as stretching, the stimulus was repeated. Crying or too active infants were given a pacifier but we tried to avoid stimulating at a time when the infant was sucking vigorously. Statistical analysis indicates that using pacifiers did not introduce a constant error.

There was a minimal interval of 15 seconds between presentations, longer if the infant remained active. The two-second duration of the stimulus application was timed by watching peripherally a timer with a large sweeping hand measuring to hundredths of a second. While this meant that the examiner had to observe both the response and the clock, it was not too difficult, and periodic two-observer checks on accuracy showed no consistent or appreciable error. In cases of uncertainty as to whether or not a delayed response had occurred within the time limit, the stimulus was repeated.

Four threshold determinations were made, using descending-ascending-descending-ascending order of presentation, and the average of these four series was taken as the threshold. In order to insure that the infant was sufficiently wide awake for test purposes, the first series was counted only after the elicitation of two consecutive responses. The first stimulus was one of medium intensity (315 volts, dial setting 5) to which most infants responded. If the infant did not respond to this first stimulus and to the one immediately below it in the series, stimuli of increasing intensity were given until two successive responses occurred. Occasionally when the ½-megohm resistance was left in, an infant did not respond at even the highest voltage available. In these cases, the resistance was taken out of the circuit and two shocks of 480 volts were administered. This is an increased amount of shock, but of an undetermined quantity since, with the resistance removed, the effect of individual skin resistance will vary from *S* to *S*. It was sufficiently intense, however, to elicit a response in all but two infants.

When the two successive responses had been obtained, the first descending series was continued with the stimulus one step below that to which the infant had responded. For example, if the infant responded to stimuli 5 and 4 (315 and 260 volts), the next stimulus was 3 (205 volts). If a response had not been obtained until stimuli 7 and 8 (in that order) were reached, the next stimulus was 7. When the ½-megohm resistance had been removed in order to obtain a response, it was replaced for the next stimulus and the series continued with the highest resistance-in stimulus (dial setting 10). The series was continued until the infant did not respond to two consecutive stimuli. The threshold was then calculated as the midpoint between the last stimulus evoking a response and the first stimulus which did not evoke a response.

There were two exceptions to the use of the midpoint as the threshold. Abnormal infants sometimes responded to 480 volts when the ½-megohm resistance was removed and failed to respond to dial setting 10 (530 volts) when it was replaced. As noted above, removal of the ½-megohm resistance increased the effective shock intensity to an undetermined extent so that we could not interpolate between these values. The threshold was therefore arbitrarily assigned its minimal possible value, that of 530 volts. A few older infants responded to dial setting 0 (50 volts). In these cases, a twelfth stimulus value (0 volts) was introduced by turning off the current. If the infant did not respond to the touch of the electrode alone, the threshold was, of course, the midpoint or 25 volts. If he did respond when there was no shock at all, the threshold (for response to shock) was zero volts.

The second series, an ascending one, began with the stimulus one step above that which completed the first series and continued through two consecutive responses. The third series, a descending series, was similarly begun with the stimulus one step below that which had completed the second series.

Occasionally, reversals occurred within a series, i.e., on a descending series, the *S*, after failing to respond to one stimulus, then responded to a less intense stimulus; or on an ascending series,

No reversal series (normal S)	Reversal series (normal S)	Reversal series (abnormal S)
5+	5+	5−
4+	4+	7−
→	rev.→	
3−	3−	9−
2−	2+	10−
3−	1−	480 v. +
→		
4+	0−	480 v. +
		→
5+	1−	10−
4+	2−	9−
3+	3−	10−
→	→	→
2−	4+	480 v. +
1−	5+	480 v. +
→		rev.→
2+	4+	10−
	→	
3+	3−	9+
	2−	8−
	etc.	7−
		8−
		etc.

→ Indicates a threshold.
+ Response.
− Failure-to-respond.

FIG. 1. Illustrations of the order of presenting pain stimuli.

after responding to one stimulus, he failed to respond to a more intense stimulus. Such a series was continued through the criterion of two consecutive responses or failures but was disregarded in threshold determinations. Instead additional series were presented until there were four nonreversed thresholds. Some illustrations of the presentation procedure are diagrammed in Figure 1.

Maturation Scale

A Maturation Scale was devised on the basis of items reported in previous studies (4, 6, 8, 10), and adapted to the needs of the present study and the nature of the present age-groups.[3] The primary criterion for inclusion of an item was that it should reflect changes with age, in this case differences between one-day-old prematures and one-day-old full term infants or between one-day and five-day-old full terms. A few items were included which did not differentiate well on the basis of age because preliminary work suggested that they might differentiate the traumatized infant. Other criteria used were that the scale as a whole should survey the total repertoire of behavior not covered in other tests of the battery, that the items should involve little subjectivity in scoring, that the total time for administration be about fifteen minutes, and that it not require any elaborate equipment.

The scale consists of nine items which receive varying credits depending upon the level of the response. If more than one kind of response occurs, the infant is given credit for the higher-scoring response. The general distinction between low-scoring and high-scoring responses is between generalized, mass movements and more specialized, stimulus-oriented responses. The maximum possible score is twenty-one.

Items do not have to be administered in any set order. Some items arouse the infant more completely than others, so that variation in order of presentation in such a way as to preserve a semiwakeful state was attempted. When crying occurred, testing was interrupted until the *S* was soothed with pacifier, rocking, or whatever method achieved results. The pacifier was not permitted during actual administration of the items. A record was kept of whether the *S*'s eyes were open or closed. If the infant was rated as not irritable (see below), was only minimally active and did not open his eyes on at least 50 per cent of the trials, conditions for testing were considered unsatisfactory. Test results under these conditions have been analyzed separately.

---

[3] The author wishes to express gratitude to Ruth G. Matarazzo for her aid in making a final revision of the Maturation Scale and for her help in clarifying instructions on this and subsequent procedures.

One to three trials were given on each item. In order to reduce testing time, if the *maximum* credit for an item was obtained on the first trial, no further trials were given. If the *same* credit was obtained on the first two trials, no further trials were given. Otherwise, *all three trials were given and the final score for the item was the average of the credits on each trial.* Generally a second trial on an item was not given immediately after a first one, but other items were interspersed.

A trial could be discarded and another trial substituted if it was felt that test conditions were unsatisfactory due to an interruption, onset of crying, etc. However, trials were generally not discarded simply because it was difficult to determine the scoring of the behavior. The practice was followed of requiring a clear-cut response in order to receive credit. In cases of indecision, credit was not given. The record blank used in scoring this test is shown in Figure 2.

Item 1. *Prone head reaction.* The infant is placed in the prone position. His arms are extended over the head with palms down and his head is adjusted so that the nose touches the bed. He is observed for 60 seconds.

One credit is given if, *within 20 seconds,* the infant turns his head sufficiently to the side so that the nose is no longer buried.

Two credits are given if, *within 60 seconds,* the infant lifts his head so that it *just clears the bed,* whether or not he first turns it. Credit is not given for clear efforts to raise the head unless the infant succeeds in raising it sufficiently so that there is a space between his face and the bed.

Item 2. *Crawls.* This item is observed during administration of Item 1.

One credit is given *if at least three alternating movements of the leg,* simulating crawling, occur in succession. Flexion of both legs followed by extension of both legs is not credited.

Item 3. *Pushes feet.* With the infant lying on his back, the examiner places one hand flat against the bottom of both feet in such a way that the infant's knees are flexed. Gentle pressure is exerted against the feet for 20 seconds.

One credit is given if there is a definite increase in the muscle tone of the legs such that the examiner can feel the increased pressure against his hand and can observe the tightening of the leg muscles.

Two credits are given if the infant extends his legs and thus pushes the examiner's hand *or,* with the legs remaining flexed, pushes against the hand with sufficient force to move himself backward. Occasionally, a Moro or a general stretching movement occurs during the administration of this item. The trial should then be repeated.

Item 4. *Auditory reaction.* With the infant lying on his back, a rattle is shaken gently several times at a distance of about six inches from either side of the head and in line with the ear. The rattle should not be in the infant's field of vision. Examiner shakes the rattle for approximately 5 seconds and then pauses for 5 seconds. This procedure is repeated for four presentations of the rattle, but with each presentation the intensity of the shaking is increased. The range is roughly from a very slight, minimal rattle to as abrupt and sharp a rattle as can be given without at the same time shaking the table or infant. After a pause of 10 to 20 seconds, a bell is sounded. This procedure was adopted because the increasing intensity of stimulation normally elicited more than one kind of behavior and made it easier to distinguish the different kinds of response scored. While more precise control of the stimulus might be desirable, it was felt that the present procedure satisfied the aims set up for the Maturation Scale. Not only would precise control of the auditory stimulus require elaborate apparatus, but there is perhaps more likelihood of eliciting several kinds of response when the quality of the sound is varied by using both rattle and bell and using noise rather than pure tones.

One credit is given for any change in behavior which is clearly in response to the bell, i.e., which occurs within 2 seconds after the bell is sounded. A Moro or partial Moro is a common response. Eye blinking or tightening of the eye muscles is a common minimal response and is given credit.

Two credits are given if the infant responds to the rattle. The same criteria are applied as for the bell.

Three credits are given if the infant appears to listen to the rattle. Opening of the eyes and decreased movement or, in older infants, head turning and eye movements, whether or not in the direction of the stimulus, are credited. This should not be scored unless it is a clear-cut response.

Trial	1	2	3
1. Head Reaction (prone position) 60"			
1 Turns head to side in 20".	--	--	--
2 Head clears bed in 60".	--	--	--
2. Crawls (prone position) 60"			
1 Three alternating leg movements simulating crawling.	--	--	--
3. Pushes feet (supine) -- gentle pressure exerted for 20".			
1 Increase in muscle tone of legs.	--	--	--
2 Legs extend, or body moves backward.	--	--	--
4. Auditory Reaction (supine) -- 5" S periods.			
1 Response to bell (startle, blink).	--	--	--
2 Response to rattle (as above).	--	--	--
3 Listening to rattle (opening eyes, decreased movement, head turning).	--	--	--
4 Persistent listening to rattle (3 of 4 presentations), or listening also to bell.	--	--	--
5 Both persistent listening to rattle and listening to bell.	--	--	--
5. Cotton (supine) 20"			
1 Any movement within 2".	--	--	--
2 Specific movements of either head or mouth (back and forth movements or sustained head retraction or movements of mouth in direction of S).	--	--	--
3 Both specific head and mouth movements present more than 50% of R time.	--	--	--
4 Head, mouth, and coordinated arm movements all present.	--	--	--
6. Paper (supine) 20"			
1 Same as Item 5.	--	--	--
2 Same as Item 5.	--	--	--
3 Same as Item 5.	--	--	--
4 Same as Item 5.	--	--	--
7. Persistence -- percentage of time infant responds to S in Items 5 and 6.			
1 Persistent R in 2/3 of trials.	--	--	--
1/2 Persistent R in 1/2 of trials.	--	--	--
8. Vigor -- rate and extent of movement in 5 and 6.			
1 Vigorous R in 2/3 of trials.	--	--	--
1/2 Vigorous R in 1/2 of trials.	--	--	--
9. Grasp (supine).			
1 Infant registers pull of more than 30 ounces.	--	--	--

TOTAL SCORE: ____________

FIG. 2. Maturation Scale record blank.

Four credits are given if either *(a)* the listening to the rattle is persistent through three of the four presentations of the rattle or *(b)* if the listening response occurs only once to the rattle but also occurs to the bell, even if followed after a brief quieting by a typical Moro response.

Five credits are given if both persistent listening to the rattle and listening to the bell occur.

Item 5. *Cotton*. With the infant lying on his back, a small piece of cotton is lightly placed so that it covers the nostrils and barely touches the upper lip. It is held in position for 20 seconds.

One credit is given for any change in behavior which is clearly a response to the stimulus. Any easily discriminated movement is credited if it follows within two seconds the application of the stimulus. Responses with a longer latency are credited if the examiner feels they are not spontaneous. Long latencies are more likely to be found in relatively inactive infants and therefore, if a fairly vigorous response involving the extremities and at least one head turn occur at any time during the 20-second presentation period (and there is no other stimulation), it may be credited. If only small movements or twitches occur, they are generally spontaneous.

Two credits are given if the response includes specific movements, of *either the head or mouth.* Credit for specific head movement is given for (*a*) several back and forth movements or (*b*) a sustained head retraction which is usually a coordinated body movement of avoidance and involves arching of the back. Specific mouth movement is scored when there is sustained opening of the mouth with movements of the lips upward in the direction of the stimulus. Sucking, yawning, or small lip movements with the mouth closed are not credited. To receive credit at this level, the specific movements should be present a majority of the time that any response is present, i.e., the characteristic mode of responding should include specific oriented movements. If the typical response is a generalized mass movement with an occasional specific response, credit is not given at this level.

Three credits are given if *both* specific head and mouth movements are present more than 50 per cent of the time the infant is responding. Criteria for a specific response are the same as above.

Four credits are given if head, mouth, and coordinated arm movements are all present. Coordinated arm movements are scored if both hands are brought to the midline more than once or are kept at the midline, and if the hands are open. The infant in these cases appears to be batting at the stimulus. A single such coordinated movement is not credited, but there may be some alternation in the kind of specific response, as long as the characteristic mode of responding is with specific oriented movements. Coordinated arm movements without head and mouth movements did not occur in our sample. Had they occurred, three credits would have been given.

Item 6. *Paper.* With the infant lying on his back, a piece of cellophane, 1½" by 2", is lightly held with one hand in such a way that it covers both the nostrils and the mouth. It is held in position for 20 seconds.

Zero to 4 credits are assigned in the same manner as for Item 5.

Item 7. *Persistence.* During each of the trials for Items 5 and 6, the examiner notes the percentage of time during which the infant is responding to the stimulus. As we found it somewhat difficult to observe the response closely and also follow a stop watch, this judgment is not a precise one. The infant is credited as persistent on a trial if he responds throughout the 20-second period. If there is a period as long as 3 or 4 seconds during which he does not respond, whether initially, during, or near the end of stimulation, persistence is not credited. The number of trials on which this judgment is made will vary from two to six, depending on how many are necessary to score Items 5 and 6. Persistence is not scored as either present or absent on a trial when there is no response to the stimulus. However, if there is no response on all 6 trials, the persistence credit is, of course, zero.

One credit is given if a persistent response occurs in two-thirds of the trials.

One-half credit is given if a persistent response is obtained in one-half of the trials, *or* if the examiner is in doubt as to whether the behavior on the majority of the individual trials should or should not be called persistent. Half-credits are allowed on this item since it requires a more subjective rating than most.

Item 8. *Vigor.* The procedure for this item is the same as for Item 7. This also requires a relatively subjective rating and half-credits are therefore permitted. The criterion of vigor was the behavior of the normal, full-term, one-day-old infant. Prematures were generally given half-credit on this item. Rate and extent of movement primarily determine the judgment of vigor, and duration the judgment of persistence. Persistent and vigorous behavior tend to occur together but may occur independently.

One credit is given if a vigorous response occurs in at least two-thirds of the trials.

One-half credit is given if a vigorous response occurs in one-half of the trials *or* if the examiner is in doubt as to whether the behavior on the majority of the individual trials should or should not be called vigorous.

Item 9. *Grasp.* With the infant lying on his back, a rubber-covered metal stirrup attached to a small spring balance is placed in the infant's palm. The apparatus and procedure are adapted from Sherman *et al.* (14). The stirrup is rubbed against the palm until the infant grasps it. A gentle, steady pull is then exerted until the infant either releases his grasp or the scale registers a pull greater than 30 ounces.

One credit is given if the strength of pull reaches or exceeds 30 ounces.

Vision Scale

Clinical eye examinations of the newborn are generally not extensive, although abnormalities such as ptosis, strabismus, nystagmus, immobility, and randomness of eye movements may be recorded when they are striking. Because these visual abnormalities are reported to occur frequently in normal newborns (5, p. 958; 12, p. 77), they are not considered of much value in diagnosis. No effort is ordinarily made to observe the eye's response to a stimulus—that is, to note the ability to fixate or pursue a bright light or similar object—as widespread medical opinion holds that these capacities are not present until several weeks after birth.

It has been reported, however, that the newborn, and even the late-stage fetus, can fixate briefly and can pursue a stimulus through an arc of 30 degrees (1; 7, p. 76 ff.). We were able to corroborate this readily. It is, of course, true that the newborn differs markedly in these respects from the two-month old. No special effort is required to elicit visual response from the older infant. His facial expression and the smoothness of his eye movements are sufficiently like those of an adult so that even the casual observer feels convinced that the infant "sees." With the newborn, on the other hand, patience and careful procedure are necessary. In the first place, it is difficult to obtain a satisfactory observation period when the child is neither asleep nor crying. Further, fixation and pursuit, when they do occur, are less clearly defined and of shorter duration. Thus, careful objective criteria for judging the response must be specified.

Despite these obstacles, it seemed worthwhile to include observation of visual performance in our test battery. A number of writers (7, p. 248; 15, p. 168) have felt that reading problems or defects in other complex visual perceptions may occur as a result of brain injuries sustained at birth. Certainly visual impairment is a common consequence of brain injury and probably one which is frequently undetected. In many cases, there is no structural change observable in the peripheral organ, and the impairment is demonstrated only through use of sensitive tests. Psychologists have found, for example, that brain-damaged individuals with normal vision may be less able to fuse flickering lights (11) or may have difficulty with such a complex visual-motor task as copying designs (9). It is possible that brain injuries at birth may not only produce undetected visual dysfunction but that this may more readily occur precisely because visual functions are in the process of developing and are therefore more susceptible to injury.

The general procedure we followed, that of observing and categorizing the response to a moving stimulus, is similar to the procedure used by Gesell (7) and adopted by other designers of infant tests.

Observations can be made only when the infant's eyes are open for a sufficient length of time. With most *Ss* this was not a major problem. They were under examination for half an hour to an hour, and during this time usually opened their eyes on several occasions. Other procedures were interrupted to take advantage of the opportunity, and the observations were repeated if the initial observation period had been too short or was otherwise unsatisfactory. If an *S* did not spontaneously open his eyes, we attempted to arouse him by various kinds of manipulation, of which the most successful was to hold him vertically either upright or upside down. If it was impossible to obtain a satisfactory observation period, the test could not be included. An observation was considered satisfactory if at least five trials had been given or an *S* had obtained a score of six in fewer trials.

The stimulus is presented with the infant in a supine position. Either the examiner's hand, a bell, or a metal tape measure 1 and 3/8 inches in diameter was used. No further effort was made to standardize the stimulus, since we were attempting to determine whether the capacity was present under the most favorable conditions. As some infants seemed to respond more readily to one stimulus than to another, the examiner used whichever one was preferred. Stimuli were generally changed after one or two trials, as there appeared to be a change in their effectiveness after repeated use.

The newborn usually does not turn head and eyes toward a stimulus at the periphery. Fixation is limited to a relatively narrow range directly in front of the eyes, an area which is determined by the tonic-neck reflex. Stimuli are therefore initially presented in this area. The experimenter observes the direction in which one or both eyes appear to be turned and places the stimulus in line with the eyes. He then moves it slowly toward and away from the infant, since the distance from the eyes at which an object can be fixated is also limited, varying from about 3 inches to 12 inches (7, p. 191). It may be necessary to shift the stimulus laterally in order to locate the place of regard. If fixation does not occur, the examiner attempts to elicit it by adding an auditory stimulus temporarily (e.g., the bell may be rung softly, the fingers snapped, etc.).

After preliminary efforts to locate the position which is optimal for fixation, there are presented a number of trials in which the stimulus is moved slowly upward from the line of regard or in either horizontal direction for as great a distance as the infant's eyes will follow. Each trial is begun with the stimulus in the place of regard. Visual items are scored both for the type of response which can be elicited and for ease of elicitation. As many trials may be given as are necessary to arrive at these judgments. Undoubtedly this procedure introduces a subjective element. However, it appears to be more adequate than specifying an arbitrary number of trials, since so many difficult-to-define factors may make one trial unsatisfactory—the infant may grow sleepy, may be startled by an extraneous noise, may fixate on something else, may begin to hiccup, etc.

Several aspects of the visual response are considered in making a classification: (*a*) presence or absence of a kind of response (such as fixation or pursuit); (*b*) the ease of eliciting the response; (*c*) the direction of eye movement (horizontal or vertical); (*d*) the distance the eye moves. These characteristics of the response were ordered according to difficulty in such a way that success on one item, such as horizontal pursuit through 30 degrees with difficulty, in most cases automatically credited the *S* with success on lower items. Likewise, failure on one item meant failure on higher items. The ordering is partly on an a priori basis—i.e., an *S* credited with "easily-elicited fixation response" is assumed to have passed "difficult-to-elicit fixation response." Other items, as whether "short vertical pursuit" is easier or more difficult than "horizontal pursuit through 90 degrees" were ordered on the basis of the preliminary findings, but credit was given for each only when the response could be elicited. Each response (or item), with the exception of the zero point, is given an arbitrary weight of one, so that the *number of items passed* is the *S*'s score.

Credit 0. There is no response of fixation or pursuit *and* one of the following abnormal features is observed: (*a*) wandering, uncoordinated movements (the eyes move constantly and generally independently); (*b*) immobilization or staring; (*c*) pinpoint pupils, although the eyes are spontaneously opened and remain open for more than a few seconds. The zero score is given only when these abnormal features are the characteristic response during observation rather than an isolated episode of nystagmus, strabismus, etc.

Item 1. No fixation or pursuit but none of the above abnormalities is observed.

Item 2. Fixation is brief or obtained with difficulty. The criteria of fixation have been summarized by Gesell (7, p. 191) as follows: "Quieting of body activity and of fussing; raising of upper eyelids and eyebrows; widening of eye slits, wrinkling of forehead, alignment of eye or eyes with stimulus, and immobilization of one eye or both." The criteria which we considered essential were those relating to the position of the eye or eyes, i.e., alignment with the stimulus and immobilization. Various other postural accompaniments, as noted above, generally occur and probably contribute to the certainty of the subjective judgment.

Item 3. Fixation is clearly present and easily elicited, i.e., occurs in a majority of the trials.

Item 4. Horizontal pursuit is questionably present. Pursuit is considered to be present when one or both eyes, having fixated a stimulus, move in the same direction as the moving stimulus. If the reaction time is short and the infant's eye follows the stimulus relatively smoothly and moves at about the same rate as the stimulus, there is little difficulty in identifying this as a pursuit movement. In some cases, however, repeated trials elicit delayed eye movements in the direction of the stimulus which do not appear to be random movements but which are not clearly pursuit movement. The present classification includes such cases.

Item 5. Horizontal pursuit movements through an arc of 30 degrees are obtained with difficulty. This classification does not indicate uncertainty as to whether or not there is a pursuit movement, but rather whether the movement could be elicited in fewer than 50 per cent of the trials.

Item 6. Horizontal pursuit movements through an arc of 30 degrees are easily obtained, that is, are present in more than 50 per cent of the trials.

Item 7. Vertical pursuit movements are obtained. These may be very short, but there is usually little problem in deciding whether they are present. Those movements of the eye which are upward random deviations generally begin with the eye in a peripheral position from which it rolls slowly upward. They are readily distinguished from the rapid vertical movements which follow fixation and may occur in pursuit of the stimulus.

Item 8. Horizontal pursuit movements through an arc of 90 degrees are obtained.

Item 9. Vertical pursuit is obtained through an arc of more than 30 degrees.

Item 10. Horizontal pursuit movements through an arc of more than 90 degrees are obtained. To accomplish this, the infant must move the head as well as the eyes.

Irritability and Muscle-Tension Ratings

In pediatric discussions of the brain-injured newborn, there are frequent references to such symptoms as a high-pitched and feeble cry, excessive irritability, muscular rigidity or flaccidity, and poor muscle tonus. We attempted to provide a crude quantification of these characteristics by means of two ratings—a rating of irritability and one of muscle tension or tone. The ratings are based on observations of the infant while other tests are being administered, and on a few simple supplementary procedures.

The experimenters had had considerable experience with the phenomena to be rated, and the ratings probably depend to a considerable extent upon "expertness." However, they have several advantages over the alternative of using reports in a medical chart which have been contributed by many different people. In addition to the immediate advantage of providing ordinal measurement rather than verbal description, finer discriminations become possible. Checks with the pediatric record showed that we did not overlook cases described in these terms but that we did detect less extreme examples.

In an attempt both to reduce the dependence on "expert judgment" and to increase reliability, we analyzed the behavior upon which such ratings would be based and broke it down into separate components. The *S*s were first rated in terms of these separate components or submeasures, and only afterward was an over-all rating made on that scale.

Irritability Rating. Irritability as used here is similar to the concept of physiological irritability—how sensitive the infant is to stimulation and not how loudly and lustily he cries. Several factors which contribute to this rating could be abstracted:

1. *The intensity of stimuli which evoke crying.* The normal infant does not usually cry when he is picked up and brought to the examining room and will be disturbed only by the more intense forms of stimulation—electrical stimulation above the threshold, loud noises, rough handling, etc. Abnormal infants, on the other hand, may squeal when first picked up and continue to fuss whenever handled or stimulated.
2. *The state of the infant.* When a normal infant cries, it appears to be because he has become too wide awake, is hungry, or is defecating. Hunger may be suspected in an infant who actively roots if the mouth is brought into contact with any surface and who vigorously sucks the offered pacifier. The normal infant also does not sustain well a too-long wakeful period. If the infant is awake when brought to examination or wakes up almost at once, he generally remains contented only for a brief period and then fairly abruptly bursts into lusty crying. In contrast, the abnormal infant, although fussing almost continuously, may remain with eyes closed.
3. *The cry of the infant.* The normal infant cries loudly and lustily. A weak, high-pitched, "fussy" cry or whimper is characteristic of the abnormal child.
4. *Ease of quieting.* The cry of the normal infant may be sustained, and he will not quiet spontaneously—suggesting that internal, not external, stimuli are responsible for the crying. The cry of the abnormal infant, on the other hand, is more likely to be intermittent and will usually cease temporarily if the infant is left alone. Occasional abnormal babies are irritable throughout but are distinguished by the characteristics noted above.

A three-point scale, with numerical values of 0, 1, and 2, was provided for the rating of irritability. It was decided in advance that the zero point should represent the behavior of a normal infant, the endpoint (value 2.0) the extreme and easily identified form of behavior seen in a grossly abnormal infant, and the midpoint (value 1.0) a "just-perceptible" form of abnormal behavior. With these points as signposts, *S*s were actually located anywhere on the continuum. Each of the points on the scale was anchored as follows:

Normal irritability. Crying, if it occurs at all, is only in response to fairly intense external stimulation or in response to fatigue, hunger, or defecation. The cry itself is loud and may be sustained.

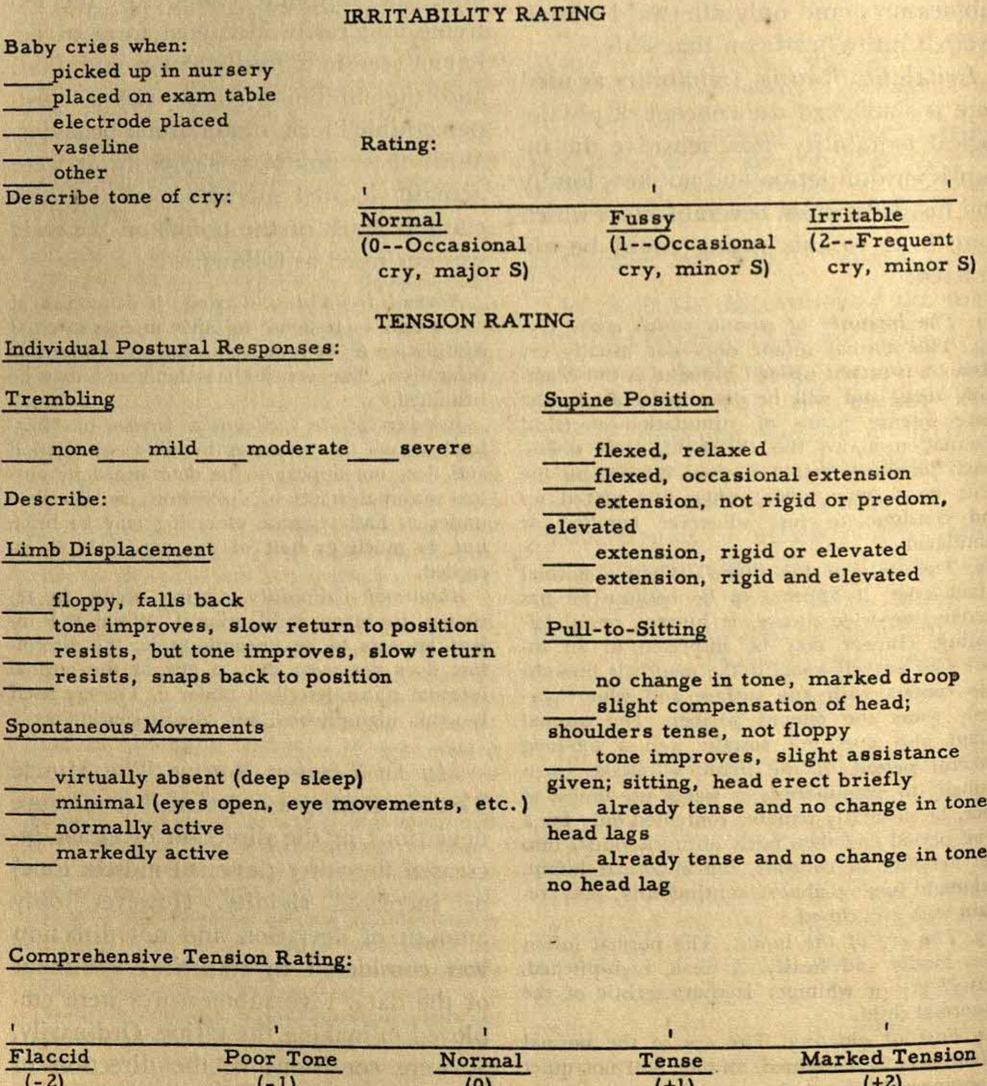
Just-perceptible irritability. Crying or "fussing" occurs in response to mild stimulation and does not appear to be determined by normal internal states of discomfort, as described under 2. Each episode of crying may be brief, but as much as half of the period is so occupied.

Abnormal irritability. Crying occurs in response to many mild stimuli. While the infant appears to be uncomfortable, the discomfort does not seem due to any of the normal internal states described under 2. The cry itself is weak, high-pitched, and almost continuous.

Muscle Tension Rating. The Muscle Tension Rating was designed to measure deviations in the direction of either increased flaccidity (lessened muscle tone) or increased rigidity. However, only amount of deviation and not direction was considered in statistical treatment of the data. Five submeasures were employed in making the rating. Ordinarily, *S*s were consistent in the direction of their response on these submeasures, but two or three exceptional cases showed mild deviations in both directions. They were rated accordingly and the deviations totaled without regard to sign.

The five submeasures required rating the following: 1, nature (flexed or extended) of the supine position which the legs assumed spontaneously; 2, resistance to limb displacement; 3, change in muscle tone in response to being pulled to a sitting posture; 4, amount of spontaneous activity; and 5, frequency of trembling of body parts and the stimuli evoking this response. The procedure for obtaining these was as follows:

1. *Leg position.* During the threshold procedure, examiner observes the leg position which was spontaneously assumed between trials. The record sheet lists five descriptive classifications of flexion-extension arranged to form a con-



IRRITABILITY RATING

Baby cries when:
____picked up in nursery
____placed on exam table
____electrode placed
____vaseline
____other
Describe tone of cry:

Rating:

Normal	Fussy	Irritable
(0--Occasional cry, major S)	(1--Occasional cry, minor S)	(2--Frequent cry, minor S)

TENSION RATING

Individual Postural Responses:

Trembling

____none____mild____moderate____severe

Describe:

Limb Displacement

____floppy, falls back
____tone improves, slow return to position
____resists, but tone improves, slow return
____resists, snaps back to position

Spontaneous Movements

____virtually absent (deep sleep)
____minimal (eyes open, eye movements, etc.)
____normally active
____markedly active

Supine Position

____flexed, relaxed
____flexed, occasional extension
____extension, not rigid or predom, elevated
____extension, rigid or elevated
____extension, rigid and elevated

Pull-to-Sitting

____no change in tone, marked droop
____slight compensation of head; shoulders tense, not floppy
____tone improves, slight assistance given; sitting, head erect briefly
____already tense and no change in tone; head lags
____already tense and no change in tone; no head lag

Comprehensive Tension Rating:

Flaccid	Poor Tone	Normal	Tense	Marked Tension
(-2)	(-1)	(0)	(+1)	(+2)

FIG. 3. Record sheet for rating irritability and muscle tension.

tinuum (see Fig. 3). After the pain threshold had been determined, a check was placed after the classification which described the position or positions assumed frequently.

2. *Resistance to limb displacement.* The examiner places his hand under the infant's heel, raises the leg through an arc of 45 degrees, and then releases it. Examiner attends both to the change in tone while lifting the leg and to the response of the leg when released. This procedure is repeated several times at intervals. The response is rated by placing a check after one of the four descriptions on the record sheet.

3. *Pull-to-sitting.* The examiner takes hold of both of the infant's hands and pulls gently two or three times without raising the child from the table, noting whether the arms respond with increased muscular tonus. Examiner then grasps the infant's forearms and pulls the body forward about 45 degrees, noting the head and shoulder response. The procedure is repeated *two or more times* before the response is rated.

4. *Spontaneous movement.* This rating is made on the basis of behavior during the entire examination. A check was placed after one of the four descriptions on the data sheet.

5. *Trembling.* Trembling of body parts is observed during the whole period of examination. A qualitative description written of such movements includes the body parts involved in such a response; how frequently it occurs; and whether it is spontaneous, in response to mild stimuli, or in response to stimuli also evoking the Moro reflex.

At the end of examination, the overall rating on muscle tension was made. A five-point scale, with numerical values from -2 to $+2$, was designed for the rating. The zero point represented the behavior of a normal infant and the endpoints (values -2 and $+2$) the extremes of flaccidity and rigidity, respectively, as seen in a grossly abnormal infant. The -1 and $+1$ points represented a "just perceptible" form of abnormal behavior. Each of the points was described in terms of the submeasures, as follows:

Abnormal rigidity. 1, Characteristic leg position is extended, rigid and elevated. 2, The legs resist displacement and snap back to the initial position when released. 3, The muscles are already tense and do not change tone in response to pull-to-sitting. There is no head lag. 4, There may be any amount of spontaneous activity. 5, Trembling of more than one body part is frequent and may be spontaneous or in response to mild stimuli as well as to startle stimuli.

Just perceptible rigidity. 1, Spontaneous leg position is extended but not rigid or elevated. 2, The legs resist displacement and snap back to the initial position when released. 3, The muscles are already tense and do not change tone in response to pull-to-sitting. There is, however, head lag. 4, There may be any amount of spontaneous activity. 5, There is usually some trembling involving more than one body part, which is pronounced in response to startle stimuli and less intense in response to other milder stimuli.

Normal muscle tone. 1, The spontaneous supine leg position is flexed, or flexed with occasional extension. 2, Muscle tone improves in response to limb displacement, although there may also be slight resistance. The limb returns relatively slowly to its initial position. 3, Muscle tone of arms and shoulders improves in response to pull-to-sitting. There is usually resistance to the pull and some head lag, but an occasional superior normal does not show head lag. These *Ss* must be distinguished from the tense infant in whom absence of head lag is not the result of improving muscle tone but of tension in the whole body so that he moves in one piece without the uneven distribution of tone characteristic of the normal newborn. 4, There may be any amount of spontaneous activity except the condition of minimal activity with eyes open. 5, Trembling, if it occurs, involves only one or two body parts, is not extensive or of long duration, and with an occasional exception, is only in response to startle stimuli.

Just perceptible flaccidity. 1, The supine leg position is always flexed. 2, There is no change in tone in response to limb displacement. The leg falls back to its initial position when released. 3, There is no change in tone in response to pull-to-sitting and there is head lag. 4, Spontaneous activity is minimal. The eyes may be open or closed. 5, There are no more than one or two instances of mild trembling.

Abnormal flaccidity. 1, Supine leg position is always flexed. 2. The initial tone of the leg is poor, feels "floppy" and weightless. There is no change in tone in response to limb displacement, and the leg falls back to its initial position when released. 3, The arm and shoulder muscle tone is initially poor and remains so even in response to pull-to-sitting. The head droops markedly. 4, Spontaneous activity is minimal even though the eyes are open. 5, There is no trembling.

SUMMARY

Five test procedures developed to provide a means of differentiating normal newborns from those who have been traumatized and possibly brain injured are here described. These were developed in the hope that differences in initial response to trauma might aid in predicting the course of subsequent development. The five tests consist of a Pain Threshold Test, a Maturation Scale, a Vision Scale, an Irritability Rating and a Muscle Tension Rating. Standardization, reliability, and differentiating ability of the procedures are presented in the second section of the study, Standardization, Reliability, and Validity (*Psychol. Monogr.*, 70, No. 21, Whole No. 428.)

REFERENCES

1. BEASLEY, W. C. An investigation of related problems in the vision of new-born infants. *Psychol. Bull.*, 1933, 30, 626. (Abstract)
2. BISHOP, G. H. Relation of pain sensory threshold to form of mechanical stimulator. *J. Neurophysiol.*, 1949, **12**, 51-57.
3. BISHOP, G. H. A simple electronic stimulator suitable for peripheral nerve and skin sensory testing. *EEG clin. Neurophysiol.*, 1953, **5**, 105-106.
4. CATTELL, PSYCHE. *The measurement of intelligence of infants and young children.* New York: Psychological Corp., 1940.
5. FORD, F. R. *Diseases of the nervous system in infancy, childhood and adolescence.* (3rd Ed.) Springfield, Ill.: Charles C Thomas, 1952.
6. GESELL, A., & ARMATRUDA, CATHERINE. *Developmental diagnosis.* New York: Hoeber, 1941.
7. GESELL, A., ILG, FRANCES L., & BULLIS, G. D. *Vision, its development in infant and child.* New York: Hoeber, 1949.
8. GILLILAND, A. R. *The Northwestern Intelligence Tests.* New York: Houghton Mifflin, 1951.
9. GRAHAM, FRANCES K., & KENDALL, BARBARA S. Performance of brain-damaged cases on a memory-for-designs test. *J. abnorm. soc. Psychol.*, 1946, **41**, 303-314.
10. GRIFFITHS, RUTH. *The abilities of babies.* New York: McGraw-Hill, 1954.
11. HALSTEAD, W. C. A power factor (P) in general intelligence: the effect of brain injuries. *J. Psychol.*, 1945, 20, 57-64.
12. PARMALEE, A. H. *Management of the newborn.* Chicago: Year Book Publishers, 1952.
13. PRATT, K. C. The neonate. In L. Carmichael (Ed.), *Manual of child psychology.* (2nd Ed.) New York: Wiley, 1954. Pp. 215-291.
14. SHERMAN, M., SHERMAN, IRENE, & FLORY, C. D. Infant behavior. *Comp. Psychol. Monogr.*, 1936, 12, No. 4.
15. STRAUSS, A. A., & LEHTINEN, LAURA E. *Psychopathology and education of the brain-injured child.* New York: Grune & Stratton, 1947.
16. WICKENS, D. D., & WICKENS, CAROL. A study of conditioning in the neonate. *J. exp. Psychol.*, 1940, **26**, 94-102.
17. WINDLE, W. F. *Asphyxia neonatorum. Its relation to the fetal blood, circulation and respiration and its effects upon the brain.* Springfield, Ill.: Charles C Thomas, 1950.

(Accepted for publication December 19, 1955)

No. 428
1956

Vol. 70
No. 21

Behavioral Differences Between Normal and Traumatized Newborns:

II. Standardization, Reliability, and Validity

By

Frances K. Graham
Washington University

Ruth G. Matarazzo
Harvard University

and

Bettye M. Caldwell
Washington University

Vol. 70, No. 21 Whole No. 428, 1956

Psychological Monographs: General and Applied

Behavioral Differences Between Normal and Traumatized Newborns:

II. Standardization, Reliability, and Validity[1]

FRANCES K. GRAHAM
School of Medicine, Washington University

RUTH G. MATARAZZO
School of Medicine, Harvard University

AND BETTYE M. CALDWELL
School of Medicine, Washington University

IN AN effort to measure behavior which would differentiate a group of infants who were "normal" from those who might be diagnosable as cases of "brain injury," the five test procedures described in the preceding paper (7) were developed. The effectiveness of these procedures is reported in the present paper.

SUBJECTS

Our subjects (*Ss*) were full term infants born on the inpatient service of the St. Louis Maternity Hospital during the period from July, 1953 through October, 1955.[2] On the basis of a sample from a private, teaching hospital, it is not possible to generalize about the frequency with which certain kinds of abnormality will occur in the population as a whole. It is probable that both socioeconomic level and techniques of prenatal and postnatal care reduced the amount of abnormality in our group. This is not a serious limitation in interpreting our results, however, since we are not concerned with absolute frequencies but with relative differences between normal and traumatized *Ss*. Whatever differences we could demonstrate between normal and traumatized groups would undoubtedly be more striking in the population as a whole where abnormality would be more marked.

The traumatized group was composed of almost the total population of traumatized infants born during the course of the study, with the exclusion of infants who were overlooked, and of a few infants whom we could not obtain permission to examine. We were informed

---

[1] This investigation is part of a long-term project being carried out in collaboration with Drs. Alexis F. Hartmann and Miriam M. Pennoyer. The work was initiated by a four-month grant from the Frances Israel Fund of the Noshin Rachmonioth Society of St. Louis, Mo. From Nov. 1, 1953 to Oct. 31, 1954 it was supported by the Children's Research Foundation, St. Louis, Mo., and since Nov. 1, 1954 by a research grant B685 from the National Institute of Neurological Diseases and Blindness, of the National Institutes of Health, Public Health Service.

[2] The writers are indebted to the authorities of the St. Louis Maternity Hospital and to the obstetricians and pediatricians who provided the opportunity for carrying out the study. We should like to express special gratitude for the cooperation of the nursing staff and especially of Miss Margaret Weber.

Bureau Ednl. Psy. Research
DAVID HARE TRAINING COLLEGE
....
Dated ....
Accs. No ....

TABLE I
CLASSIFICATION OF TRAUMATIZED SUBJECTS ACCORDING TO KIND AND DEGREE OF TRAUMA

Kind of Trauma	Degree of Trauma		
	Mild	Moderate	Severe
Anoxia	21	26	11
Mechanical trauma	0	0	3
Infections or diseases*	3	12	5
Total *N*	24	38	19

* Erythroblastosis fetalis, hypoglycemia, meningitis.

by the pediatric staff of all infants who, either at birth or subsequently, might be classified as abnormal. Infants were tested, if possible, within 24 hours after birth or as soon after that as their condition permitted. Only infants seven days old or younger are included in the study.

The kinds and degree of trauma present in the abnormal group are shown in Table 1. A pediatrician experienced in the neonatal field made the classifications without knowledge of psychological test results. A second pediatrician also rated 25 of the same cases.[3] Pediatrician One consistently rated cases more severely than Pediatrician Two, but the extent of agreement was high and significant as estimated by a correlation ratio of .86. It is important to emphasize that there is no reason to expect that all traumatized *S*s would sustain actual brain damage and that those classified as "mild" are considered by both pediatric judges to have good prognoses.

The normal group was composed of infants without prenatal, perinatal, or postnatal complications. Cases were not included where there was maternal bleeding during pregnancy or serious maternal illnesses such as rubella, diabetes, hypertension, etc. Perinatal circumstances were considered satisfactory when delivery was spontaneous or by low forceps, respiration and cry were established in a few seconds, and the infant was active. Infants were not tested on the day of circumcision or with elevated temperatures. Sufficient numbers were tested on each of the first five days of life so that the effect of age could be either statistically weighted or controlled by pairing normal with traumatized *S*s.

The primary requisite in selecting a normal control group is that it shall not differ from the experimental group except in regard to the independent variable. Unfortunately, when the independent variable is not under the control of the experimenter, there is danger that characteristics inherent in the *S* may be associated with both the independent and dependent variables and thus confound the results. There is no certain way of avoiding this, and the best that can be done is to control for such characteristics as experience indicates may be significant. In this instance, there is reason to think that prematurity, sex, race, and socioeconomic status are associated with the occurrence of trauma. However, unless they are also related to the dependent variable, i.e. test performance, the results will not be confounded and there will be no advantage to matching or pairing *S*s or otherwise controlling for possible effects.

Preliminary results indicated that,

---

[3] We are grateful to Drs Miriam Pennoyer and Marshall Greenman, of the Newborn and Premature Service of the St. Louis Maternity Hospital, for assistance in this phase of the study. Originally, they were asked to group the cases into as many discriminable ranks as possible, but their judgments were uncertain when more than three ranks were attempted. These three ranks were roughly identified as mild, moderate, and severe trauma or good, questionable, and guarded prognosis. The latter description is probably preferable and has been adopted in subsequent work.

with the exception of prematurity, these characteristics were *not* related to most of our tests. We therefore decided to test full term normal infants, without any predetermined selection criteria other than those determining classification as normal, and then to test statistically for relationships between *S* characteristics and test performance and to control such relationships as were found to exist.

The exact procedure used in selecting *Ss* was determined by the supply available in the nursery on any given day. Because of the scarcity of traumatized *Ss*, a priority system was set up. Any traumatized *Ss* available were examined first. No selection was involved here, as all such *Ss* with the exceptions noted above were examined. Next, all infants on whom blood oxygen saturation tests had been made were examined. At intervals during the period of study, blood samples were obtained from all infants born during the hours the research pediatrician was on duty. This was an unselected group of mixed normals and abnormals. All of these infants were tested at least for their pain thresholds, but the results of prematures, infants delivered without difficulty by section, and otherwise normal infants whose mothers had accidents or serious illnesses during pregnancy are not included in the present report. They do not meet the criteria for classification as either normal or traumatized. Finally, if time permitted, normal *Ss* were selected from a survey of the hospital charts. It had been planned originally to obtain all *Ss* for the normal standardization group from those on whom oxygen tests had been made. However, as there were a number of interruptions in carrying out the oxygen-testing program, it became necessary to find additional *Ss*. The psychologist looked through the hospital charts and selected the first infant who met the criteria for normality. Negro and white patients and, to some extent, private and clinic patients, were located on different floors of the hospital, so that an initial choice as regards these characteristics had always to be made. White clinic patients were most easily available to us and constitute the largest proportion of the control group. Negro and private patients were obtained only in sufficient numbers to test the relationship of these variables to test measures.

Since this was an exploratory as well as a standardization study, during which we not only were in the process of developing the techniques to be reported but also of eliminating others, not all the tests reached their final form at the same time. As soon as a method was in final form, we administered it to a sufficiently large number of normal *Ss* to provide standardization data on reliability and age changes, and it was subsequently administered only to traumatized *Ss* and those on whom oxygen tests had been made. Thus all tests were not administered to all *Ss* or to an equal number of *Ss*. Table 2 shows the distribution of socioeconomic status, sex, and race in the normal and traumatized *Ss* who were given each test, as well as in the total groups. Table 3 gives the age distribution for initial tests and retests.

PROCEDURE

The *Ss* were examined in a hospital room maintained in the same manner as the regular nursery. No soundproofing was available in this room. Whenever extraneous noises were sufficiently loud to startle an *S*, test procedures were repeated. Examinations were carried out between 10:15 A.M. and 3:00 P.M. with

TABLE 2

SOCIOECONOMIC STATUS, RACE, AND SEX OF NORMAL AND TRAUMATIZED SUBJECTS

Test	Sex		Socioeconomic Status		Race		Total *S*s
	Male	Female	Private	Clinic	Negro	White	
Pain:							
Normal	39	57	33	63	12	84	96
Traumatized	34	21	29	26	17	38	55
Maturation Scale:							
Normal	36	36	19	53	22	50	72
Traumatized	22	6	16	12	7	21	28
Vision Scale:							
Normal	40	38	28	50	16	62	78
Traumatized	23	14	20	17	10	27	37
Irritability:							
Normal	48	54	29	73	22	80	102
Traumatized	18	11	12	17	10	19	29
Tension-							
Normal	47	56	29	74	22	81	103
Traumatized	17	12	11	18	11	18	29
Total *S*s							
Normal	80	96	54	122	32	144	176
Traumatized	55	26	45	36	23	58	81

TABLE 3

AGE DISTRIBUTION OF INITIAL TESTS (I) AND RETESTS (R) GIVEN TO NORMAL AND TRAUMATIZED SUBJECTS

Test	Day 1*		Day 2		Day 3		Day 4		Day 5†		Total	
	I	R	I	R	I	R	I	R	I	R	I	R
Pain:												
Normal	61	8	9	11	8	9	10	8	8	9	96	45
Traumatized	35	1	6	11	6	13	4	7	4	14	55	46
Maturation Scale:												
Normal	24		14	7	21	4	5	17	8	2	72	30
Traumatized	11		4	1	6	5	4	2	3	5	28	13
Vision Scale:												
Normal	20		17	5	21	4	6	8	14	3	78	20
Traumatized	15		8	2	5	4	1	2	8	15	37	23
Irritability:												
Normal	31		15	6	27	5	13	12	16	5	102	28
Traumatized	16		3	2	3	5	3	3	4	3	29	13
Tension:												
Normal	31		15	6	27	5	13	12	17	5	103	28
Traumatized	17		2	2	3	4	3	2	4	4	29	12

* Includes *S*s from 12 to 36 hours old with the exception that 15 normal and 3 traumatized *S*s were given pain tests when less than 12 hours old.

† Includes traumatized *S*s from 5 to 7 days old.

most *Ss* seen during the morning. As pointed out previously, the number and kind of tests given varied during the course of the study. For those *Ss* who were tested with the final battery, the pain threshold was obtained first. Vision tests were given whenever the infant opened his eyes, and the maturation and tension scale items were given in whatever order best maintained the infant in a satisfactory state. Irritability was rated at the end of the examination. Detailed instructions for each of the procedures is given in the preceding paper (7).

RESULTS[4]

Reliability

In the present research, we are interested in three levels of measurement which require different methods of estimating reliability. In the first place, we are interested in differentiating groups. The reliability of the group measurements would be satisfactory if the error variance of the group means were significantly less than the differences between normal and traumatized groups. This will be dealt with in a later section.

Secondly and primarily, we are concerned with the accuracy of identifying individuals as members of a group. Satisfactory reliability for this purpose would be achieved if errors in measurement were not large enough to change the individual's classification as normal or abnormal.

Thirdly, we are interested in the reliability of individual raw scores, partly because there are standard methods of presenting reliability data in this form and partly because, even where the interest is centered on classifying in two categories, there is always the possibility that raw scores or transformations of them can later be shown to measure degrees of abnormality or normality. We will, therefore, present reliability measures in terms of raw scores, although raw-score reliability will underestimate the reliability of classifying as normal or abnormal. It is only for this twofold classification that validation has been attempted. Three estimates of reliability have been made: (*a*) single-session reliability, given by split-half product-moment correlations, (*b*) test-retest agreement after 24 hours, and (*c*) interscorer agreement.

1. *Split-half reliability.* Only the pain threshold score could be satisfactorily divided into two comparable halves. Split-half correlations (1st and 2nd halves) for normal and traumatized *Ss* are shown in Table 4. The correlations range from .82 to .97. Reliability is highest for the traumatized group, but this is undoubtedly due to the wider range of thresholds obtained in this group. When correlations are calculated separately for traumatized *Ss* with thresholds in the same range as the normal group, the difference in reliability disappears. Age heterogeneity apparently contributes little to raising the correlation. When age is partialled out, the correlation drops only from .87 to .82. It should

TABLE 4

SPLIT-HALF RELIABILITY OF THE PAIN THRESHOLD

Subjects	*N*	*r* (corrected)
Normal *Ss*	96	.87
Normal *Ss* (partial *r* without age covariance)	96	.82
Traumatized *Ss* (total)	55	.97
with normal scores	35	.82
with above normal scores	20	.93

---

[4] The authors are grateful to Robert C. Bilger and to John C. Glidewell for advice on statistical treatment.

be pointed out that momentary fluctuations in the state of an infant are much more marked than in an adult and a split-half reliability estimate includes correspondingly more of the possible variable error.

2. *Test-retest reliability.* Test-retest correlations are reported only for normal *S*s. Since age is a variable rather than a source of error in the present study, changes associated with it should not be included in a measure of unreliability, unless the effect of age is relatively the same for all *S*s. The effect of age is undoubtedly *not* the same for all infants in the traumatized group, as they are recovering at different rates from different kinds and degrees of trauma and there is no way of separating the differential changes due to age from those due to unreliability. Therefore test-retest correlations in this group could give little information on reliability. As far as normal *S*s are concerned, we do not know whether they change differentially with age, although recovery from even a normal birth probably varies. However, there can be no objection to regarding test-retest correlation as at least a minimum estimate of reliability.

Table 5 shows the number of *S*s who were re-examined on each procedure and the agreement of test and retest score. Product-moment correlations were calculated for Pain Threshold and for Maturation and Vision Scale scores. Percentage of perfect agreement was used in the case of the rating scales. Retest reliability is lower than split-half, as would be expected, but it is satisfactory for the kind of tests used, especially in view of the age changes which are significant at this time of life.

3. *Interscorer reliability.* On all measures except the Pain Threshold, scores depend to some extent upon judgments made by the examiner. In order to determine whether such judgments constituted a major source of unreliability, two examiners simultaneously scored a sample of *S*s, drawn at intervals throughout the course of the study, to insure that agreement did not vary over a period of time. The examiner who administered the tests had no knowledge of whether *S*s were traumatized or not traumatized. Table 6 shows the number of *S*s in each group observed by two examiners and the satisfactorily high interscorer agreement. As in the case of test-retest agreement, Tension ratings appear more reliable than those for Irritability.

Objectivity

Ideally, in such a study as this, all measurements should be made without knowledge of how the *S* is classified. It was not possible to do this routinely, but there are several lines of evidence which

TABLE 5

TEST-RETEST AGREEMENT AFTER TWENTY-FOUR HOURS

Test	Normal *S*s	Agreement
Pain	24	$r = .69$
Maturation Scale	20	$r = .62$
Vision Scale	20	$r = .62$
Irritability	28	75% perfect agreement
Tension	28	86% perfect agreement

TABLE 6

INTERSCORER AGREEMENT

Test	Normal	Traumatized	Agreement
Maturation Scale	19	5	$r = .97$
Vision Scale	17	4	$r = .90$
Irritability	20	5	68% perfect agreement
Tension	20	4	79% perfect agreement

indicate that we were successful in avoiding bias.

1. While there was often knowledge of the presence or absence of trauma, whether it was mild, moderate, or severe was *not* known. As will be subsequently shown, test results not only differentiated normal from traumatized *Ss* but also were correlated with the degree of trauma.

2. Interscorer reliability was high (see Table 6) for *Ss* scored by two examiners simultaneously. The examiner who administered the tests on these check cases was without knowledge of the *S*'s classification.

3. A group of 16 traumatized and 31 normal *Ss* was examined without any knowledge of classification. The 16 traumatized *Ss* had been given a total of 41 tests and the 31 normal *Ss* a total of 68 tests. The scores of each *S* on the tests he was given were paired with the scores of an *S* tested with knowledge of classification. Pairing took into account age, race, and kind and degree of trauma (when present). The Sign Test was used to test the hypothesis that the score differences obtained under the two conditions were randomly distributed as to direction. For this analysis, scores on all tests were considered simultaneously. Results showed that there were no significant differences in either the normal or traumatized *Ss* to suggest that scores were more "normal" or more "abnormal" when there was knowledge of the classification of the infant than when there was not such knowledge. Since bias might occur more easily on some of the more subjective tests than on others, *F* tests were applied to each test separately. Of the 10 comparisons made, normal and traumatized groups being treated separately, only one indicated a difference significant at the .05 level. That was on the Tension Rating scale, where 14 normal *Ss* tested with knowledge were rated lower than those tested without knowledge.

Variables Affecting the Performance

The normal and traumatized groups differ considerably in respect to age, sex, race, and private or clinic status, as may be seen from Tables 2 and 3. Before the performance of the two groups could be compared, therefore, it was necessary to determine whether or not these characteristics were related to the scores obtained by normal *Ss*. The existence of a relationship in traumatized *Ss* which did not exist in normal *Ss* would not confound the results but would be assumed to be an interaction effect dependent upon trauma.

In order to determine the effect of so many variables without allowing concomitant variation, it was necessary to select smaller samples from the pool of *Ss* in such a way that factors other than the one being tested were held constant. Male and female *Ss*, paired for race, socioeconomic status, and age, did not show differences on any of the tests. Private and clinic *Ss*, paired for race and age but with sex allowed to vary since it did not influence performance, also did not differ from one another on any of the tests. The effects of race, age, retesting and condition of the *Ss* were similarly tested. The *F* ratio was used to compare scores on the Pain Threshold test and on the Maturation and Vision Scales. Chi square was employed in comparing scores on the two rating scales. The ratings were categorized as "0" or "not 0" in order to obtain sufficiently large theoretical frequencies but, even so, the expected frequencies were less than 5 in two of the eight comparisons. However, none of the probabilities calculated from the chi-square distribution approached significance and would not do so even if calculated directly.

Private-clinic status and sex of the *S* did not measurably affect performance on any of our tests, and there was no practice or learning effect from retesting when age was held constant. However, age itself was significantly related to score on three tests. Older *Ss* were found to be more sensitive than younger to Pain, and to perform better on the Maturation and Vision Scales. Negro *Ss* were also superior to whites on the Maturation and Vision Scales but there was no difference on the other tests. Unless the condition of *Ss* being given the Maturation Scale was "satisfactory," as previously defined, significantly poorer scores were obtained. This factor was not important, in so far as it was measured, on other tests. Where we have made assertions that a factor did affect performance, the usual 5 or 1 per cent levels of significance of difference are to be understood. Where the statement is made that a factor did *not* affect performance, the *F* ratios were in every case smaller than 1.0, and the highest chi-square probability was .18. There was, therefore, nothing to suggest that an effect would be demonstrated if another sample or a larger one were obtained.

Differences Between Normal and Traumatized Groups

Mean difference. In order to compare the performance of normal and traumatized groups on the test battery, it was necessary to equate the groups for those variables which have been shown to be related to test score. As neither the irritability nor tension ratings were affected by the variables tested, no pairing was necessary on these measures. The entire sample of normal and traumatized *S*s could be used, therefore, with the exception of those rated on irritability a day or two after circumcision. Statistical analysis, not reported, showed that this factor did raise the irritability rating. Table 7 shows the size of the groups, the variables on which they were equated by pairing, the mean scores of the two groups, the statistic used in estimating probabilities, and the probability that differences between groups are due to chance. The means are included on all five measures, although they were of course not used when chi square was the comparison statistic. On both the pain thresholds and the Vision Scale there was a significant difference in the variance of the traumatized and normal groups. In evaluating the mean difference, therefore, t was calculated according to the Cochran-Cox method with no assumptions about variance. On all tests the performance of the traumatized groups was significantly poorer than that of the normal groups.

Shape of the distributions. In the preceding section, evidence was presented that the normal and traumatized groups differed significantly, but nothing was said about the shape of the distributions. Figure 1 shows the distribution of pain thresholds in the two groups. Percentage frequency rather than frequency is plotted on the ordinate to equate the size of the groups. Standard scores are plotted on the abscissa so that thresholds of individuals tested on different days could be combined. Since the standard score transformation is a linear transformation, the shape of the curve is altered only to the extent that the distributions of the several days show differences in skewness. Standard scores of traumatized *S*s are based on the normal group.

TABLE 7

COMPARISON OF NORMAL AND TRAUMATIZED GROUPS ON FIVE TESTS

Test	N	Variables Controlled	Mean Scores	Comparison Statistic	p
Pain Threshold					
Normal	55	Age	165	t test	.01
Traumatized	55		270		
Maturation Scale					
Normal	28	Age, race	13.0	F test	.05
Traumatized	28		10.6		
Vision Scale					
Normal	37	Age, race	6.8	t test	.01
Traumatized	37		4.2		
Irritability					
Normal	91	——	.12	Chi square	.01
Traumatized	29		.61		
Tension					
Normal	103	——	.08	Chi square	.01
Traumatized	29		.48		

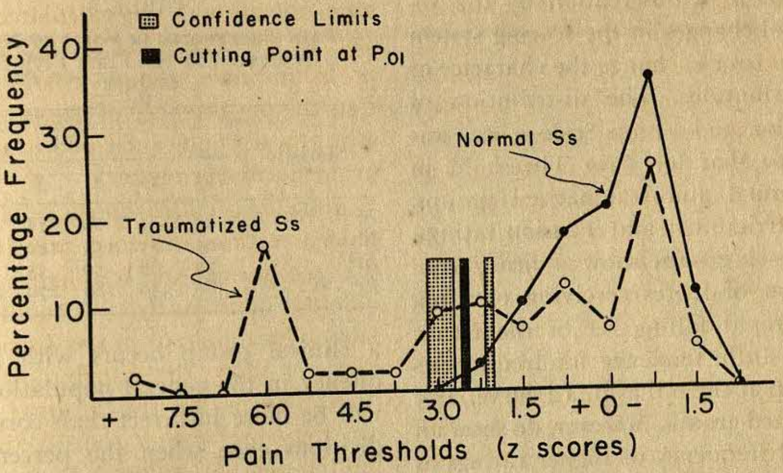


FIG. 1. Pain thresholds (in standard score form) of normal and traumatized subjects.

The graph shows a skewed distribution in the normal group, the skewness probably reflecting the effect of a lower limit to the pain threshold. Thresholds of the traumatized group, on the other hand, cover a much wider range and do not appear to fall into any type of unimodal distribution. Unless it is assumed that the considerable variability has distorted what would otherwise be a unimodal distribution, we must assume a bi- or multimodal distribution. Since the tendency to bimodality is also present in each of the three subsamples of mild, moderate, and severely traumatized infants, the hypothesis of bimodality is strengthened.

If thresholds of traumatized *Ss* are not unimodally distributed, what does this signify? It can only mean that pain sensitivity is not affected in a unitary fashion under conditions of trauma. More than one factor must be present. It is possible that the presence of a new factor is due to the greater intensity of stimulation used with some of the traumatized *Ss*, rather than to the trauma per se. This would be the case if other sense modalities were activited when a certain intensity of stimulation is reached. It is tempting, however, to speculate that the new factor is related to brain functioning. Is there a threshold for impairment of brain functioning such that on one side of the threshold, the same factors determining pain sensitivity in normal brains are operating while, once beyond the threshold, sensitivity is determined by changed condition of the brain? It is idle to speculate on the brain physiology which might be involved; there are many physiological phenomena which show this all-or-none character. For present purposes, the point is of interest in selecting cutting scores and making predictions. If the nonunitary character of pain sensitivity is due to changes in brain functioning, we would expect to find that *Ss* who later show evidence of brain damage would be selected from among those who form a second mode and not from among those who fall at the upper end of the normal distribution.

Graphic distributions of the other four tests are not presented. Interpretation of them is complicated by the fact that scoring was empirically determined on the basis of observation by the authors, and changes in the scoring system would, of course, change the character of the distributions. The distribution of Maturation and Vision Scale scores was similar to that for Pain Threshold in both normal and traumatized groups. On the Irritability and Tension ratings, the normal groups show a heavy concentration of scores receiving 0 rating with a rapid falling off of the curve. There is little tendency for frequencies to pile up at the tail as in a J curve. The traumatized groups, however, do show an increased frequency of higher ratings as well as a wider range.

Cutting Points and Normative Data

In order to identify those *Ss* among whom we expect to find later evidence of brain damage, it is desirable to establish a cutting point. It would be possible to correlate scores on each of the tests with later results. However, since we are dealing with a phenomenon of low frequency in the total population, it is likely that the percentage of "hits" can be increased by restricting ourselves to a more eligible subsample. If the bimodal distribution of scores in the traumatized group is related to brain function, it becomes even more important to separate the population forming the second mode from that forming the first mode, since there would be no reason to expect differences in scores around the normal mode to be correlated with later damage.

TABLE 8

PAIN THRESHOLDS IN VOLTS OF NORMAL SUBJECTS FOR THE FIRST FIVE DAYS OF LIFE

Statistic	Day					
	.25	1	2	3	4	5
N	15	54	20	17	18	17
Mean	205	185	155	140	120	98
SD	70	68	50	55	35	35
$p = .01$	380	350	280	270	195	183

It is also desirable to have a cutting point which will include a minimum of false positives, i.e., normal *Ss* incorrectly called abnormal. From the application of Bayes's theorem, which Meehl has made (12), it is easily demonstrated that when a clinical group occurs with low frequency in the general population, there will be more incorrect than correct predicitions even when the percentage of true positives approaches 100, unless false positives are kept at a minimum.

These considerations led us to select a cutting point at the extreme of the normal distribution—that point below which only 1 per cent of the normal population would fall. Cutting points for pain thresholds, Maturation, and Vision scores were set at t values with probabilities at the .02 level (or .01 level for a single tail), rather than at the observed $p = .01$ value in the sample. Basing cutting points on t values, when justified by an approximately normal distribution, appears preferable to using sample percentile points. Separate cutting points were determined for each day, and, in the case of Maturation and Vision scores, for both Negroes and whites. These cutting points as well as the means and standard deviation are shown in Tables 8, 9, and 10.[5] The reliability of the cutting points for pain thresholds can be estimated from the 95 per cent confidence limits, which are shown in Fig. 1. Bars rather than lines

---

[5] Because race and age did affect performance they required separate norms. An additional 89 *Ss* were therefore tested in order to increase the size of N on the Maturation, Vision, Irritability, and Tension scales.

TABLE 9

MATURATION SCALE SCORES OF NORMAL SUBJECTS FOR THE FIRST FIVE DAYS OF LIFE

Statistic	Day 1	2	3	4	5
White *S*s:					
N	37	21	20	21	20
Mean	11.2	11.9	12.8	13.5	14.1
SD	4.3	2.8	2.3	2.9	2.2
$p = .01$	.8	3.9	7.1	6.3	8.6
Negro *S*s:					
N	28		20		
Mean	12.4		13.7		
SD	3.0		2.6		
$p = .01$	4.9		7.0		

identify the upper and lower confidence limits since this is a composite curve of the samples on each of the five days. The limits are sufficiently narrow so that false positives would vary only between 0 and 3 per cent if the cut is located at any point within the interval. The number of true positives could vary more widely, but, even at the upper limits, none of the cases at the second mode is excluded. The confidence intervals for cutting points on the Vision Scale are also narrow. False positives would vary only from 0 to 2 per cent and true positives from 38 to 43 per cent. The Maturation Scale shows less discrimination than the other measures to begin with. While changes in the location of the cut would not increase the number of false positives, the discrimination of true positives could be cut from 25 to 7 per cent.

TABLE 10

VISION SCORES OF NORMAL SUBJECTS FOR THE FIRST FIVE DAYS OF LIFE

Statistic	Day 1	2	3	4	5
White *S*s:					
N	37	26	23	20	24
Mean	4.9	6.5	6.4	6.1	6.9
SD	2.2	1.3	1.4	1.5	1.9
$p = .01$	0	3.4	2.8	2.3	2.3
Negro *S*s:					
N	23		21		
Mean	5.4		7.6		
SD	2.0		2.0		
$p = .01$	.5		2.6		

In determining cutting points for the Irritability and Tension Scales, it was necessary to use the observed *p* values since the distributions depart radically from the normal curve as can be seen from Table 11. However, ties in score (at a rating of 1.0) occurred in the lower 6 per cent of Irritability Ratings and 3 per cent of Tension Ratings so that a value unique to the first percentile could not be determined. While it is possible to divide the tied cases into appropriate proportions above and below the cutting point, this is not satisfactory when a decision must be made about classifying an individual *S*. The cut must be placed either just below or just above the score. The decision was made to place it above the rating of 1.0 on the Irritability Scale, thus excluding 99.5 per cent of normal *S*s, and below the 1.0 rating on the Tension Scale, which excluded only 96.8 per cent of normals. These decisions took into account the shape of both the normal and traumatized group distributions and may to some extent have capitalized on chance fluctuations. Estimates

TABLE 11

PERCENTAGE OF NORMAL SUBJECTS RECEIVING A GIVEN IRRITABILITY OR TENSION RATING SCORE ($N = 186$)

Scale	Rating Score 0	.25	.5	.75	1.0	1.25 or above
Irritability	76.3	5.9	10.2	1.1	5.9	0.5
Tension	79.1	8.6	7.5	1.1	2.7	0.5

of the reliability of these cuts are necessarily crude. If 95 per cent confidence limits based on the binomial distribution are determined, we can expect the percentage of false positives on the Irritability Scale to vary from 0 to 4 per cent and the percentage of false positives on the Tension Scale to vary from 1 to 7 per cent.

These cutting points are only tentatively identified. The optimal cut, in the sense of giving minimal overlap in both directions, lies at the intersection of the normal and traumatized group distributions. For the reasons outlined above, we did not feel that such a cut would be useful in a situation where base rates are presumably low and where it is more important to exclude false positives than to reduce false negatives. Final decisions about the value of these cutting points depends upon their relationship to signs of brain damage in the *S*s' subsequent development. For this reason, and because the points are determined entirely by the normal samples, which are reasonably large and reliably measured, we did not reserve a portion of the data for cross validation.

A rough measure of the relative discriminating power of the five tests may be obtained by comparing the percentage of traumatized *S*s called abnormal. Table 12 shows the percentage of *S*s who score on the abnormal side of the cutting point on any one or more tests and on each test separately. When *S*s were retested, the poorest performance has been taken as the score on a test. These data are supplied for the normal and traumatized groups and for the three subsamples of traumatized *S*s. Pain thresholds and the Vision Scale are superior, but all tests identify some *S*s as abnormal. The percentage identified as abnormal appears to increase with the degree of trauma and, if scores on all tests are considered, is statistically significant at the .01 level when tested by chi square.

Intercorrelations

How are the various tests related to one another? Since they were designed initially to detect differences between traumatized and normal *S*s rather than among normals, the intercorrelations might be expected to differ for the two kinds of *S*s. They are therefore presented separately, as well as for the total group, in Table 13. As pointed out earlier, all tests were not given to all *S*s. The traumatized cases included in the correlations are all those *S*s who had

TABLE 12

PERCENTAGE OF SUBJECTS IDENTIFIED AS ABNORMAL BY SCORES BELOW THE CUTTING POINT ON THE DAY OF POOREST PERFORMANCE

Test	Normal	Total Traumatized	Mild Trauma	Moderate Trauma	Severe Trauma
Pain	1	42	43	30	57
Maturation Scale	0	25	(33)*	12	(50)
Vision Scale	1	41	(17)	31	60
Irritability	1	28	(33)	8	46
Tension	3	34	(33)	23	46
Any one or more tests	4	51	46	37	84

*Percentages in parentheses are based on an *N* of less than ten.

TABLE 13

INTERCORRELATIONS AMONG TEST PROCEDURES

Test	Normal *Ss*		Traumatized *Ss*		Total group	
	N	*r*	*N*	*r*	*N*	*r*
Pain-Mat.	46	.14	26	−.04	72	.06
Pain-Vision	46	−.11	40	−.03	86	.06
Pain-Irrit.	46	.43**	42	.08	88	.07
Pain-Tens.	46	.03	41	.16	87	.21*
Mat.-Vision	46	.38**	22	.31	68	.36**
Mat.-Irrit.	46	.09	27	−.02	73	.12
Mat.-Tens.	46	.01	26	.20	72	.20
Vision-Irrit.	46	−.24	30	.43*	76	.49**
Vision-Tens.	46	−.17	30	.15	76	.13
Irrit.-Tens.	46	.21	42	.30*	88	.39**

scores on at least two tests. The 46 normal cases are the total number of *Ss* to whom all five tests were given. Standard scores were used and the direction of scores changed so that all correlations may be similarly interpreted.

Table 13 shows that the intercorrelations are, in fact, different for the two groups. The Maturation and Vision Scales, both presumably measuring developmental level and having considerable spread in both groups, are positively correlated with one another, although the correlation is not statistically significant in the traumatized group. As would be expected, Irritability and Tension are positively correlated in the traumatized group but not in the normal. They are measures designed to detect deviations in the direction of abnormality and therefore have little variability in the normal group. Irritability, a measure of abnormality, and Vision, a measure of developmental level, are significantly related only in the traumatized subjects.

The finding of a correlation among normal *Ss* between Pain, which does have spread within the normal range, and Irritability, which does not, is somewhat surprising. One can only suggest that the irritable infant is responding to many stimuli in a diffuse fashion and therefore tends to be less sensitive or less set for responding to a specific stimulus in a specific way. This same analysis might be expected to hold true among traumatized *Ss* except for the fact that some traumatized infants become more irritable while others become obtunded. The obtunded infant is scored as normal in Irritability but will, of course, be relatively insensitive to pain as well as to other stimuli.

Perhaps the most significant finding is that all intercorrelations are low. This is understandable since trauma to the newborn may be manifested in a variety of ways, some of which are incompatible with one another, as, for example, obtundity and hyperirritability. Such relatively low intercorrelations among the different tests, together with the adequate discriminating ability of each considered separately, points to the advisability of using a combined score on all procedures as an impairment index. This would achieve one of the characteristics desirable in constructing a test battery—i.e., low intercorrelations among the tests and high correlations with the criterion.

Another possible way of improving predictive accuracy would be to consider the length of time that an *S*'s performance remains abnormal. In retesting traumatized *S*s, we observed that the length of time scores remained abnormal varied considerably from infant to infant. The duration of an abnormal performance thus provides another dimension along which to measure the infant's response to trauma and offers promise as an additional way of identifying those *S*s on whom trauma will leave a permanent imprint.

DISCUSSION

The group of tests we have used samples much of the repertoire of an infant's response to his environment. All of the responses are relatively simple, but they represent a substantial portion of the most complicated behavior which an infant of this age can show. How complicated is such behavior? With the exception of the two rating scales, the tests can be described as measuring sensorimotor ability, i.e., (*a*) the capacity to respond at all to various kinds of sensory stimuli, and (*b*) the extent to which the response is specific to a particular stimulus. The ratings of irritability and of muscular tension provide two more dimensions along which all responses of an infant, both spontaneous and elicited, may be described. We should like to know whether measuring such behavior gives any information about either past or future development.

Does sensorimotor functioning reflect the developmental level of the nervous system? Gesell (6) has carried out extensive studies of premature and full term newborns which suggest that simple sensorimotor abilities are in the process of development during the last months of fetal life and therefore should provide measures of the stage of development of the nervous system. This pioneer work needs to be confirmed, however, by studies using standardized procedures whose reliability can be established. The difficulty we experienced in adapting from Gesell a reliable maturational scale and the widely divergent placement of similar items on various infant scales indicate that slight variations in method can produce large differences in results.

Does sensorimotor functioning at birth predict later development? Efforts to answer this question have extended over several decades, but it is still not possible to formulate definite conclusions. This is partly because similar data have been interpreted by different standards. A correlation of a given size does not arouse the same response in all psychologists. The data, however, are also conflicting. Some studies have shown small negative correlations between performance in early infancy and several years later (10, p. 637f). On the other hand, sizable positive correlations have been reported between tests at six months and tests as late as three years (14). Such factors as inadequately standardized tests, difficulties in controlling the infant's momentary state (3), differences in the kind of infant tests, and differences in the kind of development which is being predicted all contribute to the confusion.

Irwin (9) has questioned the reasonableness of expecting to predict "intelligent" behavior from neonatal response under even the most favorable conditions and with sensitive tests. Neurophysiological work suggests that the newborn is not capable of cortical functioning (4, p. 60f). While there is opposition to this view, the argument is mainly whether there is no cortical activity or some minimal amount (11). Irwin's (9) position is essentially that you cannot expect to predict the future functioning of the cerebral cortex on the basis of tests made at a time when the cortex is *nonfunctional.* On logical grounds alone, this is not a necessary conclusion. There is no a priori reason why the functioning of subcortical structures should be uncorrelated with later cortical development. From Hebb's (8, p. 109ff) interesting hypothesis that the higher organism is the slower learner initially, one might even infer that there is a negative correlation. Practically, however, the greater the gap between phenomena observed and those to be predicted, the more difficult it is likely to be to establish relationships.

The present work is not primarily concerned with predicting the relative superiority of "normal" individuals, but rather in determining whether external trauma has caused brain injury. We did find that a considerable percentage of traumatized infants show impaired functioning as compared with nontraumatized newborns, and that such impairment is related to clinical judgments of severity of trauma. But will measures of impairment of a newborn predict the extent of later impairment? The question cannot be answered at the present time. One can say only that it seems reasonable to assume that the greater the present trauma, the greater the likelihood that some cells will suffer irreversible damage. Since the cerebral cortex of the infant is relatively nonfunctional, is it likely that trauma to the cortex could be detected by newborn tests? The answer would seem to be no, if the cortex is, in fact, completely nonfunctional and if the trauma were limited only to the cortex. Destruction of a part of the nervous system before it is mature enough to function will not give rise to the expected symptoms until "the time arrives for that structure to play its proper role" (4, p. 69). However, many kinds of trauma, and especially anoxia, may produce diffuse, multiple lesions in both cortical and subcortical structures (1). With such trauma, the extent of disruption of subcortical functions might provide an index of the degree of total damage.

We suggested that a general impairment index could be computed by combining the number of tests showing impairment and the length of time impairment persists. The extent to which such an index will be valuable depends on whether or not specific areas of damage are associated with impairment on specific tests. Without discarding the possibility of specificity entirely, we would relegate it to a minor role. Functioning of a newborn is relatively undifferentiated and it seems likely that the extent of trauma would be the most important factor in determining subsequent impairment. This could be true either if sizable injuries in any area affect most functions, or if there is a tendency for the kinds of trauma to which a newborn is exposed to cause injury in the same areas because these are more susceptible to insult.

It would be interesting to know more about the nature of the impairment that occurs. Because the intensity of the pain stimulus was systematically varied, we were able to observe that in infants whose pain thresholds were abnormally high, the ability to perceive the stimulus did *not* seem impaired, since many *Ss* cried or gave other general responses to weak stimuli. Similarly, motor pathways were apparently intact, since the required leg movements could occur independently of stimulation. The difficulty, then, must lie in a failure to integrate stimulus and response in such a way that a response appropriate or specific to the stimulus could occur. In the development of the nervous system, sensory and motor nerves are in contact with their respective organs before it is possible for an excitation to pass from sensory to motor mechanism. We may speculate that, in the loss of the specific connection between stimulus and response under conditions of trauma, we see an example of the phenomenon observed in adult brain-damaged individuals, namely, that those abilities most recently acquired are most readily lost.

One other phenomenon deserves mention. Some of the traumatized infants might be described as hyperreactive with increased muscular tension and irritability. These infants were sensitive to any kind of mild stimulation but gave generalized rather than specific responses. Other traumatized infants showed diminished general activity, were flaccid and apathetic. Deviations in either direction could impair performance on various tests. The direction of deviation was not taken directly into account in the scoring, although apathetic infants were more likely to be penalized on the Maturation Scale. Are these two kinds of impairment related to the severity of the trauma? Is it possible that they are analogous to stages in the development of coma in adults? Initially, "in anesthesia and in coma induced by changes in the internal environment (decrease in the oxygen tension or in the blood sugar level) the cortex and brain stem suffer opposite changes in excitability" (5, p. 212). While cortical activity is diminished, there is increased reactivity of the brain stem. As coma deepens, brain-stem activity likewise decreases (5, p. 217).

We have been interested in differentiating those infants whose trauma is contemporaneous with the examination. What can be said of the possibility of detecting injury which has occurred some time prior to birth? This will depend on the extent to which our measures reflect temporary effects due to a present disturbance of the nervous system. The newborn traumatized early in intrauterine life should not appear irritable or obtunded, and if the damage suffered has been restricted to the cortex there might be no behavioral manifestations at birth. On the other hand, if subcortical centers have been damaged, we might expect impairment of sensorimotor functioning. We had the opportunity of testing three mongols—one of the few kinds of fetal injury which, because of the physical peculiarities, can be detected at birth. All three did show impairment on Vision and Maturation Scales and also showed the marked muscular flaccidity which characterizes this condition. Their pain thresholds, however, were normal and, except for one infant with an intestinal infection, they were not irritable.

In concluding, an encouraging word ought to be said for the value of infants as psychological *Ss*. In the last decade, they have been largely overlooked, yet their very youngness offers unique advantages in psychological areas where the complexity of social and other environmental influences makes it difficult to disentangle the relevant variables. There is ample room for improvement in techniques of measurement. Our relatively crude efforts at objectification have been rewarding and have suggested many lines which might profitably be followed with more precise methods.

We aimed to develop a short battery of tests requiring minimal equipment in order that a large number of *Ss* could be studied quickly. Because of these requirements, some promising methods were dropped which ought to be investigated more fully. With photographic equipment, for example, thresholds could be obtained for the pupillary response to light. Electronic equipment is now available that would permit measurement of several aspects of infant activity level. Early work by Richter (15) and Wenger (17) related the height of the skin resistance level to states of tension and wakefulness. Individual differences in hydration and in toughness of the skin present serious problems, but the desirability of obtaining an independent and objective measure of the state of an infant is great.

The most conspicuous omission from our battery is a measure of learning. A series of early studies attempting to establish classical conditioning in the newborn were disappointing (13, p. 376f) but these studies did not exhaust the possibilities of studying the conditions for and kinds of changes in behavior which newborns can exhibit. We found that, except for the rating scales, performance on our tests improved with age during the first five days of life. In the case of pain thresholds, the improvement was linearly related to age change, as can be seen in Figure 2. Is this improvement due to maturation, to experience, to recovery from the trauma of birth, or to a combination of these factors?

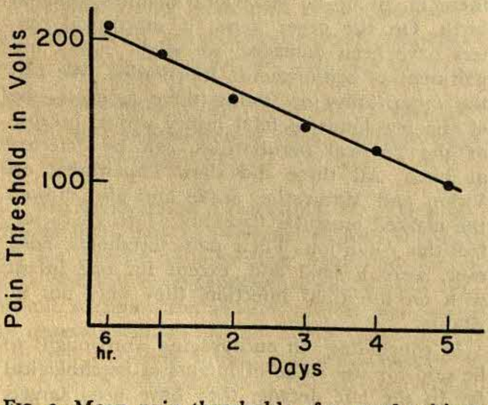


FIG. 2. Mean pain thresholds of normal subjects on the first five days of life.

Without raising the question of whether the improvement is due to maturation or learning, one may ask what it is that has improved. In the case of the pain thresholds, for example, has the infant become more sensitive to pain as he grows older, or does the age change represent increased ability to give a differentiated response? Our results cannot separate the two possibilities. It would be interesting to determine whether the age changes are still present when any kind of response to the stimulus is recorded, i.e., when a generalized rather than a differentiated response is measured. If there are *no* age changes in the ability to make a generalized response, then the hypothesis of increased sensitivity could be rejected. This procedure was actually used in two previous studies, one by Dockeray (2) in 1934, and one by Sherman, Sherman, and Flory (16) in 1936. Unfortunately, their results disagree, which is perhaps to be expected since only very crude control of stimulus intensity was possible at that time. Repetition of this experiment with modern methods of stimulus control should be able to provide a more conclusive answer.

SUMMARY

Five test procedures, described in the preceding paper (7), were administered to 265 infants without prenatal, perinatal, or postnatal complications and to 81 infants suffering from anoxia, mechanical birth injury, or diseases or infections associated with brain damage. The traumatized newborns composed nearly the total population of such infants born at the St. Louis Maternity Hospital during a two-year period.

The five tests consist of a Pain Threshold Test, a Maturation Scale, a Vision Scale, an Irritability Rating and a Muscle Tension Rating. Reliability of the procedures was variously measured by split-half correlation, test-retest agreement, and interscorer agreement as applicable. All of the tests appeared to be satisfactorily reliable. A sample of 109 test scores obtained without knowledge of *Ss'* classification did not differ significantly from those obtained under the usual conditions of partial knowledge.

Norms have been presented for each test, with separate norms provided for each of the first five days of life and for Negro and white *S*s where these variables were related to performance. Older *S*s were found to be more sensitive than younger on the Pain Threshold Test and to perform better on the Maturation and Vision Scales. Negro *S*s were superior to whites on both the Maturation and Vision Scales, but there was no race difference on the other tests. Private-clinic status and sex of the *S* did not measurably affect performance. There was no practice or learning effect from retesting when age was held constant.

Normal and traumatized groups, paired for relevant variables, obtained significantly different scores on all tests. When a cutting point at the poorer extreme of the normal distribution was selected, all tests identified some traumatized *S*s as abnormal while false positives ranged only from 1 to 3 per cent. The percentage identified as abnormal increased with the seriousness, as rated by pediatric judges, of the trauma. Since intercorrelations of the five tests were low, a combined abnormality score was tentatively recommended.

The question of whether newborn behavior can predict either past or future development of the infant was discussed. Cautious consideration was also given to the relationship between the present findings and neuro-physiological knowledge about functioning of the newborn brain.

REFERENCES

1. Courville, C. B. *Cerebral anoxia.* Los Angeles: San Lucas, 1953.
2. Dockeray, F. C., & Rice, Charlotte. Response of newborn infants to pain stimulation. *Ohio State Univer. Studies. Contrib. Psychol.,* 1934, No. 12, 82-93.
3. Escalona, Sibylle. The use of infant tests for predictive purposes. *Bull. Menninger Clin.,* 1950, **14,** 117-128.
4. Ford, F. R. *Diseases of the nervous system in infancy, childhood and adolescence.* (3rd Ed.) Springfield, Ill.: Charles C Thomas, 1952.
5. Gellhorn, E. *Physiological foundations of neurology and psychiatry.* Minneapolis: Univer. of Minnesota Press, 1953.
6. Gesell, A., & Armatruda, Catherine. *Developmental diagnosis.* New York: Hoeber, 1941.
7. Graham, Frances K. Behavioral differences beween normal and traumatized newborns. I. The test procedures. *Psychol. Monogr.,* 1956, **70,** No. 20 (Whole No. 427).
8. Hebb, D. O. *The organization of behavior.* New York: Wiley, 1949.
9. Irwin, O. C. Can infants have IQ's. *Psychol. Rev.,* 1942, **49,** 69-79.
10. Jones, H. E. The environment and mental development. In L. Carmichael (Ed.) *Manual of child psychology.* (2nd Ed.) New York: Wiley, 1954.
11. Kleitman, N. The role of the cerebral cortex in the development and maintenance of consciousness. In H. A. Abramson (Ed.) *Problems of consciousness.* New York: Josiah Macy Jr. Foundation, 1955.
12. Meehl, P. E., & Rosen, A. Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. *Psychol. Bull.,* 1955, **52,** 194-216.
13. Munn, N. L. Learning in children. In L. Carmichael (Ed.) *Manual of child psychology.* (2nd Ed.) New York: Wiley, 1954.
14. Nelson, Virginia L., & Richards, T. W. Studies in mental development. I. Performance on Gesell items at six months and its predictive value for performance on mental tests at two and three years. *J. genet. Psychol.,* 1938, **52,** 303-325.
15. Richter, C. High electrical resistance of the skin of newborn infants and its significance. *Amer. J. Dis. Children,* 1930, **40,** 18-26.
16. Sherman, M., Sherman, Irene, & Flory, C. D. Infant behavior. *Comp. Psychol. Monogr.,* 1936, **12,** No. 4.
17. Wenger, M. A., & Irwin, O. C. Fluctuations in skin resistance of infants and adults and their relation to muscular processes. *Univer. Iowa Stud. Child Welf.,* 1936, **12,** 141-179.

(Accepted for publication December 19, 1955)

Vol. 70, No. 22 Whole No. 429, 1956

Psychological Monographs: General and Applied

The Sense of Effectiveness in Public Affairs*

ELIZABETH DOUVAN
Survey Research Center, University of Michigan

AND ALAN M. WALKER
Earlham College

I. THE PROBLEM

A. INTRODUCTION

THE ubiquitous phenomenon of the nonparticipating citizen has become a concern, not only to political scientists and political theorists, but to social scientists across discipline lines. It has become increasingly well recognized that the behavior of that large portion of our population which does not vote and takes no interest in public affairs requires for its analysis a multivariable scheme. While it is useful to inspect particular elections for conditions which may account for lack of interest, it is also fruitful to approach the problem at a psychological level and inquire into the motivational and attitudinal structure which distinguishes the interested from the apathetic citizen.

A variable that has attracted considerable interest and has proved useful in efforts of the latter kind is an attitude we have called the "sense of effectiveness"—that is, the feeling or sense of being an effective agent in the world of public affairs. One group of authors, speaking of a "sense of political efficacy," define it as "the feeling that individual political action does have or can have an impact upon the political process . . . that political and social change is possible and that the individual citizen can play a part in bringing about this change" (3, p. 187). It is this psychological variable and its determinants with which the present analysis is concerned.

A part of the analysis from the present study, which is reported elsewhere (4), was designed to test the hypothesis that people who feel effective will be characterized by more highly differentiated attitudes and a livelier interest in particular public issues than will people who feel unable to have any influence in the political sphere. Consistent attitudinal differences were found among groups based on a measure of effectiveness.

In the analysis presented in the present report, we have focused on the question of the position of the sense of effectiveness in the individual's psychological structure. In the first section the question posed is that of the generality of

---

* The study reported here was done in the Public Affairs Program of the Survey Research Center, of which Stephen B. Withey is Program Director. Professor Angus Campbell, the Center's director, Benjamin J. Darsky, and Professor Max L. Hutt all made major contributions to the design of the study. Support for the research was granted by the Phoenix Project of the University of Michigan, a memorial fund dedicated to research on peacetime implications of atomic energy. We are deeply grateful for Phoenix support, without which the study would not have been possible.

effectiveness as a psychological characteristic; in the second we present an analysis of certain projective materials in an effort to clarify the relationship of this attitudinal variable to two ego-level personality factors.

B. Methods of the Present Investigation

1. Subjects and Procedure

Data for the analysis were derived from a personal interview survey conducted with 316 members of the labor-force population of Detroit. The complex area sample design customarily employed by the Survey Research Center was the basis for the original selection of respondents.[1] The interview for the present study was administered only to members of the sample who were employed, seeking employment, or only temporarily unemployed.

Interviewing was done in respondents' homes by trained interviewers. A standard questionnaire was used, and in most cases the questions were of an "open-ended" type permitting full and free responses.

2. Measures

a. *Control procedures.* Riesman and Glazer (9) have defined apathy in psychological terms—and have embraced within the concept the two variables of affect and competence. They maintain that competence alone—that is, the feeling that one can have an effect on political events, without an accompanying attachment to the world of public affairs in the form of a lively interest in or concern about it—is not enough to distinguish a person as having an effective orientation toward political events.

In the present study we were concerned primarily with the sense of effectiveness—a variable similar to the "competence" which these authors treat. The variable of affective involvement, however, entered our study design as a control for the following reason.

The problem of enlisting the motivation of subjects in an experiment or study is common to many areas of psychological research.[2] We might anticipate that lack of involvement toward the task would be a particularly grave and common source of error in studies of attitudes toward objects or issues remote from subjects' lives. Such questions, if asked in the abstract, would have relatively little meaning to many people, and would elicit little psychological involvement.

We therefore asked the questions on citizen influence—designed to measure the sense of effectiveness—in the context of two problems of broad national significance—the threat of atomic war and the problem of inflation.

Following this discussion, we put to respondents the questions on citizen influence. These were phrased in general terms about the influence a citizen could have on decisions like *these* (relating to the problems—e.g., what steps to take to prevent war, whether to build an Hbomb, how to control inflation).[3] The emphasis in the discussion of problems was on the threat of atomic war, and government decisions demanded by this threat. More questions were asked about this problem than about the economic issue. It is also a more affect-laden issue for most people, and did elicit more emotionally colored responses. It was this issue which set the tone of the discussion.

---

[1] A detailed description of this sampling method is presented by Kish (6).

[2] The significance of this problem of enlistment of involvement for the general questions of reliability and validity of measures is particularly great in attitude research. When we make predictions about relationships among attitudinal variables, we in a sense assume that motivation toward the task of responding to question-measures is relatively constant for all subjects. Either a lack of motivation toward the task (wherein responses to various questions may be superficial or dictated by fluctuating whim) or the presence of extraneous motivation (e.g., the desire to avoid revealing one's true attitude) may create error in the test of an essentially correct hypothesis.

If in response to the questions on war threat there was an indication that the respondent felt no involvement in the issue—because he felt the possibility of war an unreal threat—he was dropped from the analysis of effectiveness. Since he did not accept the reality of the problem context of the interview, it was felt that his answers to questions about the influence people could have on such issues would not be based on a suitable level of involvement. A subsidiary analysis was made on these cases, and is reported elsewhere (4).

Responses to the following question and probe were inspected in this control procedure.

> a. Would you say you're very concerned about it (the threat of war), not at all concerned, or what?
>
> b. How do you mean?

Two groups of respondents were eliminated on this basis: those referred to in the previous paragraph, who were not concerned because they did not consider the problem real—they did not think there was any threat of war (48 cases); and those who were not concerned because they had such faith in the nation's leadership that they felt all would be well (12 cases).

b. *The sense of effectiveness.* The remaining cases were grouped into three categories of felt effectiveness in the following way: a direct question was asked about the extent to which the average citizen can influence government decisions:[4] "How much of an influence do you think the average citizen can have on decisions like these?" Respondents were divided into two groups on the basis of responses to this question—those who thought that they could have "quite a lot" or "some" power, and those who felt they could have "very little" or "none!" A sizable number of respondents mentioned spontaneously, when the problem of war was discussed, that they felt they personally could have no effect

TABLE I

EFFECTIVENESS CATEGORIES AND THEIR FREQUENCIES

Category and Definition	Number of Cases
1. High Effectiveness Feel that the average citizen can have "quite a lot" or "some" influence on government decisions	91
2. Medium Effectiveness Feel that the average citizen can have "very little" or "no" influence	93
3. Low Effectiveness Feel the average citizen can have "little" or "no" influence, and raise their own ineffectiveness in earlier discussion of issues	60
Total	244

---

[3] In the political study previously referred to (3), the measure of political efficacy did not immediately follow discussion of particular issues. In that case, however, the entire questionnaire was devoted to an imminent election and the issues surrounding the campaign, so that involvement and context were already established.

[4] We have assumed that when an individual answers a question about the influence an *average citizen* can have on government decisions, he is, in virtually all cases, answering on the basis of the degree of influence he feels that he personally can have.

on public affairs. Since the questions on problems preceded the discussion of citizen influence, and contained no reference to personal influence, we considered spontaneous mention of personal ineffectiveness a sign of the salience of this attitude in the psychological structure of the individual. For people who interject the note of helplessness when they are initially confronted with a public issue, this psychological characteristic must be an important part of the view of the broader world.[5] The third—and least effective—group was formed, then, of those respondents who thought that the average citizen could have little or no influence, and who also raised their own sense of ineffectiveness spontaneously in the earlier discussion.

The groups that emerged from this categorization, with their frequencies, appear in Table 1.

In addition to respondents who were dropped from the analysis because they failed to meet our criterion of involvement, three cases were dropped because they raised their feeling of ineffectiveness in the discussion of war, but contradicted this position on the later direct question about the average citizen's influence. Ten cases were dropped because they gave no response to one or another of the criterion questions.

---

[5] While we recognized that war is a unique public issue and that individuals who raise their own helplessness with respect to war might not if the issue were less extreme, we nevertheless maintain the point that people who interject personal weakness in this context are likely to be more dominated by a sense of powerlessness generally than those who do not raise the issue.

II. THE QUESTION OF GENERALITY: THE SENSE OF EFFECTIVENESS IN RELATION TO MORE PERSONAL ATTITUDES

In a great deal of the writing on the sense of effectiveness the assumption is made that this attitude in relation to public affairs is the same as a more personal sense of competence or potency, or at any rate that it is simply a social manifestation of a more pervasive personality characteristic. It seems to the present authors that this assumption requires testing.

In the present section of the analysis, then, we have posed the question of the generality of the sense of effectiveness. Is the individual who feels that he can have a measure of influence on public affairs, more than the ineffective citizen, characterized by a feeling of competence or potency in other areas of life? Put another way, the question is whether the sense of political effectiveness stems primarily from experiences in public affairs and is therefore relatively uncorrelated with any similar sense of competence in interpersonal or occupational (or other more day-to-day) situations—or whether, on the other hand, it is a symptom, in one area, of a more general characteristic of the person's conception of the world, which will be manifest in other areas as well. Our general prediction was that the latter view would prove closer to the case.

We predicted that people who had a sense of effectiveness in public affairs would be characterized by greater satisfaction, security, and feelings of competence in their personal lives than those who felt ineffective. We speculated, with previous writers, that "people who feel impotent, dependent, or completely occupied with personal problems will not have the energy and courage necessary" to maintain the kind of commitment to broad issues implied in a high degree of the sense of political effectiveness (7, p. 65). Although there are obviously other possible bases for an active involvement in public affairs—e.g., some people may compensate for personal insecurities through an active and effective role in civic affairs—we predicted that in general a positive relationship would exist between feelings of personal competence and the sense of political effectiveness.

A. SATISFACTION, SECURITY, AND COMPETENCE IN THE PERSONAL SPHERE

For measuring the more personal variables, we included questions about the respondent's own life and his feelings about his personal and interpersonal affairs. The questions, listed below, can be grouped psychologically into three classes: those dealing with the respondent's satisfaction or dissatisfaction with his present life; those dealing with the individual's expectation of positive or negative outcomes in the future; and those which probe the respondent's perception of himself as an effective or noneffective agent in the determination of his own future.

The questions, grouped under appropriate headings, read:

Satisfaction

a. Some people feel that their lives have worked out just the way they wanted. Others feel they've really had bad breaks. How do *you* feel about the way your life is turning out?

b. How does your life compare with what you wanted it to be like?

c. Some people feel quite secure financially; others have many worries about how they will get along. How is it in your case?

d. If you had a choice, would you like to see a son of yours do the same kind of work you do, or some other kind of work? (If different) What kind?

Security About the Future

e. What do you think your chances are of living the kind of life you'd like to have? Do you think they are pretty good, not so good or what?

f. What do you think your chances are of being promoted in the job you have now?

g. On the whole, what would you say your future looks like at the place you work—pretty bright, not much to look forward to, or what?

h. Would you say your job is a steady and secure one, or one that could easily fold up?

Perception of Self

i. Some people feel they can make pretty definite plans for their lives for the next few years. Others feel they aren't in a position to plan ahead. How about you—do you feel able to plan ahead or not?

j. When you *do* make plans, do you usually feel they're almost certain to work out, or that you can't count on them working out, or what?

To form a more reliable measure of feelings about personal affairs, the data from Questions a, e, and i in the preceding list—each a generally phrased question from one of the three areas—were combined in an "index of personal competence."

In addition to the theoretical ground for combining these items—i.e., that they represented three important elements of personal security—the combination yielded an index of reproducibility of .94. This indication of a relatively high degree of internal consistency provided a methodological justification for considering the combined items a single index. Nonscalar groups were collapsed into the scale patterns having the same number of positive responses.

The major scale patterns are as follows:

a. *High.* People who give confident answers to all three questions.

b. *Moderately high.* Individuals who are satisfied, feel optimistic about the future, but do not feel that they are in a position to plan their own futures.

c. *Moderately low.* Individuals who are satisfied with their lives, but feel their chances for attaining their aspirations are poor and that they are not in control of their futures.

d. *Low.* Those who are dissatisfied, feel their chances are poor, and that they have no control over the future.

The relationship between the effectiveness variable and personal reactions was tested with the index of personal competence and also with responses to a number of individual items. Throughout this part of the analysis, chi square was used to test differences among groups, and product-moment coefficients were computed to obtain estimates of correlation. The correlation coefficients were derived by the method suggested by Wert *et al.* (10, pp. 305-313) for broad categories, and were corrected for coarse groupings.

B. Results

The central notion of this section of our analysis was that, in general, individuals who feel effective in the public sphere will be characterized by greater feelings of security in the personal environment as well. We predicted that those who feel unable to operate meaningfully in the public sphere would, more often than not, manifest a sense of being thwarted, of apprehension and lack of control with respect to events of everyday living.

Before presenting the tests of this hypothesis, a word should be said about the role of demographic characteristics in the determination of the sense of effectiveness. In research dealing with social attitudes, it is characteristic to find substantial relationships between psychological variables and education, income, and/or occupation. With an attitude like the sense of effectiveness, there is good reason to anticipate such findings. Assuming that actual power to affect public decisions is differentially distributed in the various strata of society, then indices of status like education or income should relate to *feelings* of effectiveness as well.

In other studies of the sense of effectiveness (2,3), these relationships have occurred, and we had planned in the analysis of the present study to control for critical demographic characteristics. The preliminary analysis indicated, however, that our measure of effectiveness did not relate significantly to status measures. The correlation between education and effectiveness was .11; between income and effectiveness was .09; and between occupation and effectiveness was .09. Based on chi square, the probabilities of such relations arising by chance were, respectively, between .10 and .20, $>$ .80, and $>$ .20,

There are two possible explanations for this difference between our results and those of previous analyses. The first is that our measure is not identical to those used in other research. Our inclusion of spontaneous references to respondents' lack of influence may change our measure enough to explain the unique absence of relations to demographic factors. We also phrased questions in terms of the average citizen's, rather than the respondent's own, influence. Another possible explanation however, lies in the nature of the population we sampled.

Our respondents were all drawn from the labor force of Detroit. This selection tended to reduce the range of demographic variables. Detroit is a high wage-rate city, and, in addition, our exclusion of the permanently unemployed or retired from our sample meant that we interviewed no people of extremely low income, and may have excluded some of the very high as well.[6] By reducing the range of a variable, we obtain a measure of control over it, and we also reduce the chances of significant relationships occurring when the variable is run against any other.

The fact that demographic factors were not significantly related to the effectiveness measure simplified the analysis, since no special control on them was necessary in testing relationships between attitudinal variables.

---

[6] While it is not possible to make a definitive test of the extent to which the nature of our sample accounted for the absence of relationship between status and the sense of effectiveness, we were able to demonstrate that this factor had some influence. Using as a base, data from the political study conducted by the Survey Research Center (3), we selected from a national sample all interviews taken in metropolitan areas with employed persons. In this group, compared to the full national sample, there was a reduction in the relationship between status and "political efficacy." The range in status measures was, even in this case, larger than the range for our Detroit labor-force sample.

1. *Effectiveness and the Index of Personal Competence*

The relation between the two major measures is presented as item 1 in Table 2, The results bear out our prediction. The highly effective group of respondents scores significantly higher on the competence index than does the low group, with the medium group about in the middle.

Similar tests of the hypothesis with responses to individual questions follow under the aspect of competence to which they are appropriate.

2. *Current Satisfaction or Feeling of Being Thwarted*

A number of comparisons of our effectiveness groups were made in order to test

TABLE 2

RELATION BETWEEN SENSE OF EFFECTIVENESS IN PUBLIC AFFAIRS AND ATTITUDES TOWARD EVENTS IN THE PERSONAL SPHERE (SHOWING PROPORTIONS GIVING SELECTED RESPONSES)

Personal Attitude[a]	Effectiveness			χ^2	p	r
	High ($N=99$)	Medium ($N=93$)	Low ($N=60$)			
1. Index of personal competence						
a. High	.45	.30	.22	24.6	<.001	.47
b. Low	.02	.15	.27			
2. Outcome of life plans						
a. About as planned	.51	.54	.35	3.65	<.10>.05	.17
3. Satisfaction with outcome of life plans[b]	.82	.73	.59	6.6	<.05	.28
4. Feeling of financial security[c]	.68	.65	.41	8.6	<.01	.26
5. Attitude toward occupation						
a. Want son to follow similar occupation	.23	.11	.06	9.8	<.01	.30
b. Want son to follow different (unspecified) occupation	.07	.12	.23			
6. Expectation of promotion[d]						
a. Chances are good	.43	.30	.07	14.1	<.001	.38
7. Evaluation of job future[d]						
a. Positive evaluation	.73	.65	.52	4.1	<.10>.05	.25
8. Feeling of job security[e]	.85	.84	.74	1.7	<.30>.20	.11
9. Index of job outlook[d]						
a. High positive index	.30	.21	.05	16.2	<.001	.43
10. Security about personal plans[f]	.85	.68	.54	24.8	<.001	.37

[a] For verbatim phrasing of the measures in this column see text, Section I.

[b] The number of cases is reduced since it was not always possible to tell whether the respondent was satisfied or dissatisfied with the comparison. Therefore, not all respondents could be rated. The numbers should read: High 55, Medium 55, Low 34.

[c] The number of cases was reduced because Form II of the questionnaire did not include the question on financial security, and unemployed respondents were not asked the question. The numbers should read: High 66, Medium 61, Low 44.

[d] The number of cases is reduced by the exclusion of the unemployed, self-employed, and those given Form II of the questionnaire. The numbers should read: High 60, Medium 56, Low 41.

[e] The number of cases is reduced because unemployed and self-employed people were not asked this question. The numbers should read: High 77, Medium 74, Low 56.

[f] The number of cases is reduced here for the following reason. During the early stage of the interviewing there was some confusion on the part of a few interviewers about the question on plans working out, with the result that they were using it as though it were contingent on the person's saying in the previous question that he felt in a position to plan ahead. Actually the question was meant to be asked of all respondents. Those cases where it was not asked were eliminated from the table. The numbers should read: High 83, Medium 84, Low 52.

the hypothesis that ineffectiveness is related to feelings of frustration and of being thwarted in the personal sphere.

Respondents were asked "How does your life compare with what you wanted it to be like?" Results on this question are in the direction predicted and are consistent with the other findings of this section. They appear on line 2 of Table 2.

Respondents who are *low* on effectiveness are least likely to think their lives are similar to what they had anticipated. There is, however, no substantial difference between the other two groups on this response, and the small difference which does appear is not in the direction predicted.

It was possible, of course, that a person could feel his life had turned out quite differently from what he had expected or planned on, yet be satisfied with the outcome. Coders were therefore asked to rate respondents, where possible, as satisfied or dissatisfied with the comparison described above between their plans and life situation. Results of these ratings are presented on line 3 of Table 2, and indicate that satisfaction with present life condition is related significantly to the sense of effectiveness.

Although we have seen that there is no significant relationship between income and feeling of effectiveness, we had predicted that the less effective groups would report *feeling* economically thwarted more often than the effective group. In Table 2, line 4, we see that this was in fact the case.

The fact that the ineffective groups feel thwarted in their occupational roles can be seen in responses to the question, "If you had a choice, would you like to see a son of yours do the same kind of work you do, or some other kind of work?" (If different) "What kind?" (Table 2, line 5.)

Respondents in the least effective group are least likely to say they want their sons to do work similar to their own. More often than other respondents they reject their own work as an ideal for a son, but suggest no alternative occupation. It is as though they were saying: "Let him do something else. I don't know what, but not what I'm doing."

3. *Anticipation of Future Satisfaction*

Included in the competence index was the item: "What do you think your chances are of living the kind of life you'd like to have?" Other questions concerning anticipation of future success or failure were asked in the context of occupational achievement. It should be kept in mind that our effectiveness groups did not differ significantly in occupational classification.

Results of two occupational future questions are presented on lines 6 and 7 of Table 2. The second question produced results which do not meet our test of significance, but are in the predicted direction.

The final question on occupational future was: "Would you say your job is a steady and secure one, or one that could easily fold up?" Results of this question, seen on line 8, were not significant, but again were in the direction predicted. Apparently such reality factors as seniority are operating here to reduce the relationship between outlook toward one's occupational future and personal effectiveness. There was so little distribution in responses to this question (nearly everyone feels his job is secure) that it would be unusual to find a high degree of association.

Since the occupational life of an individual is so important—simply on a time basis, it occupies a central position in adult life—we wanted to derive the most reliable possible measure of the occupational outlook to relate to effectiveness.

The three items on occupational future just presented were therefore combined into a single measure. As we would expect from the unity of content, they scaled significantly better than could be expected by chance. (Coefficient of reproducibility = .98.) Nonscalar groups were collapsed into the scale patterns having the same number of positive responses, and the two lowest scale groups were combined because of small numbers. Anticipation of positive outcome in one's job should certainly be a good indication and perhaps determiner of a feeling of effectiveness. We predicted a positive relationship between job outlook and effectiveness. The results presented as item 9 in Table 2 bear out this prediction.

4. *Control Over Future*

In addition to the question on planning included in the competence index, respondents were asked, "When you *do* make plans, do you usually feel they're almost certain to work out, or that you can't count on them working out, or what?" Again the prediction was that those people who feel ineffective with respect to national affairs would feel least able to count on personal plans working out. This, it seemed to us, would represent a general apprehension about the future and a feeling that the future was determined more by external events than by events or actions subject to the respondent's own control. In item 10 of the summary table we see that the results correspond to the prediction.

Following the question "What are your chances of living the kind of life you'd like to have?" respondents were asked why they thought their chances were good or poor. Factors mentioned were categorized according to whether they were personal (intelligence, hard work, skill) or environmental (financial situation, family situation, job situation). We predicted that less effective people, when they say their chances are poor, would attribute their poor chances to more external factors not subject to their own control. Table 3 shows the proportions of both types of factors given by our groups in cases where they say their chances are good and where they are said to be poor.

It is interesting to note that the effective group attributed their chances, whether good or poor, about equally to personal and environmental factors. (The variation in proportions does not differ from chance expectancy.) On the other hand, the low effectiveness group

TABLE 3

RELATION BETWEEN EFFECTIVENESS AND ATTRIBUTION OF GOOD OR POOR PROSPECTS TO PERSONAL AND ENVIRONMENTAL FACTORS

Good prospects attributed to	Effectiveness High	Medium	Low
	($N=75$)	($N=61$)	($N=36$)
Personal factors	47%	59%	64%
Environmental factors	49	39	36
Not ascertained	4	2	—
	100%	100%	100%
Poor prospects attributed to	**High**	**Medium**	**Low**
	($N=15$)	($N=31$)	($N=22$)
Personal factors	47%	33%	20%
Environmental factors	48	64	68
Not ascertained	5	3	12
	100%	100%	100%

stresses personal factors when success is anticipated and environmental factors when they anticipate failure. The shift in proportions of personal and environmental factors cited by this group would be expected by chance less than one time in a hundred.

C. Summary on the Question of Generality

The sense of effectiveness in response to public affairs has been shown to be related to a feeling of competence in a variety of areas in the more immediate sphere of personal affairs. The relationships found are moderate in size. The correlation coefficients range from .11 to .47, with the majority clustering around .30. The fact that the two composite measures of personal security (i.e., the index of personal competence and the index of job outlook) yield the highest correlations, seems to indicate that more refined measures might reveal a closer relationship between personal competence and the sense of political effectiveness. It is nevertheless obvious that there are other important determinants of the sense of political effectiveness with which we have not dealt in the present study.

The order of magnitude of relationships demonstrated and the consistency of these relations appear to justify the interpretation that the sense of political effectiveness is to a significant extent a reflection of a more general psychological competence. In the remaining pages of this report we will turn to the question of the relation of "competence" to certain "deeper" personality variables, in an effort to clarify further the reticulate connections between our social attitude (the sense of effectiveness) and other aspects of the individual's stable psychological structure.

III. THE SENSE OF COMPETENCE IN RELATION TO TWO EGO VARIABLES

In previous studies a passive-receptive response to an environment which is considered uncontrollable has been linked theoretically to the defense mechanisms of denial and projection. In the Authoritarianism series (1) it was demonstrated that individuals who scored high on ethnocentrism and latent fascist attitudes were generally passive in orientation toward a hostile and unpredictable world. It was also shown that this picture of a hostile world had intrapsychic origins. Authoritarians were people who could not accept normal hostility in themselves, and dealt with these ego-alien elements by denying their internal existence and projecting them into the environment.

Mussen and Wyszinski (7) found a generally external orientation among politically apathetic individuals. They theorize that political apathy is based on a defense pattern similar to that of the authoritarian personality—that is, the apathetic individual, unable to accept his own hostility, shifts it into the external world, which then becomes threatening. The difference, they speculate, between the apathetic and ethnocentric individuals lies in the means by which the two handle this self-made hostile world. The ethnocentric person imputes hostility to a restricted segment of his environment (e.g., an ethnic minority) and expresses his hostility—now legitimized by the evil nature of the enemy outgroup—with impunity. The apathetic individual, on the other hand, responds by withdrawing from a more diffusely hostile and threatening world.

We included in the design of the study a series of more or less projective instruments for the purpose of measuring two aspects of personality. These were: (*a*) the nature of the respondent's conception of the external world, and (*b*) his response to internal reality—that is, to normal impulses and feelings. The inquiry into these variables and their relations to competence was prompted by our interest in establishing more specifically the meaning of the competence variable.

Competence, as measured in this study, is a conscious-level phenomenon—a feeling of relative satisfaction with and control over one's own life. The person scoring low is one who experiences frustration and thwarting and feels that the determination of his life is somehow out of his hands.

The question that next arises concerns the source of this feeling of competence. Does it stem solely from a concept of a capricious and oppressive environment, or does it also reflect a similar relation to the internal environment which is projected onto the outer world? The question reduces, in one sense, to the amount of information revealed by the competence index—whether it indicates both the level of comfort in dealings with the environment and the nature of the process on which the level is based, or only the former.

One might almost say that the very fact that the apathetic individual is not able (because of social or personal pressures) to delineate one aspect of the environment as a receptacle and restrict his hostile projections, leaves him no alternative but to withdraw. If the whole world is threatening, there is not even a focus for retaliation.

We predicted, in line with this previous work, that highly competent individuals would, in general, have a more positive outlook toward the external world, and also that they would, on the average, be more tolerant of normal impulses and feelings than those individuals who feel less competent.

A. Measures of Response to External and Internal Reality

The following series of questions was asked of each respondent:

Now, I'd like to ask you a few questions about the way people sometimes feel, and what you think they should do about it.

1. Take a person who has an awful lot of worries, and they're really getting him down. What do you think he should do about it?
2. What do you think makes some people worry a lot when other people hardly ever worry.
3. In general, what things would you say help a person most to get ahead these days?
4. Some people feel that it's good for everyone to get really angry now and then, and "blow off steam." Others say it's childish to let yourself feel so angry. How do you feel about this?
5. We hear a lot these days about different ways to bring up children. Would you tell me whether you agree or disagree with these statements about raising children? (*Probe for each:* Why?)
   a. Children should have no worries—they should be protected from troubles as much as possible.
   b. Parents should plan out the child's program each day so they're sure he is learning the right things.
   c. Children should be taught to fight their own battles as early as possible.
   d. Parents should make children try things that are a little beyond them in order to make them ambitious.
6. Do you think it's a good idea or a bad idea to tell children that babies are brought by the stork? Why?
7. What would you say was the nicest thing about being a child?
8. If a daughter of yours were getting married, what things would you want to know about the man she was marrying? Anything else?

Responses to these questions were analyzed for indications of the respondents' conception of external reality and acceptance or rejection of normal impulses. The questions, the variables to which they refer, and the code categories used as indicators, appear in Table 4.

In addition to these questions, a three-card version of the TAT (cards 3BM, 6BM, and 20) was administered to a subsample of 76 cases, and the story responses were scored for the same variables. Scoring dimensions and critical categories are presented in Table 5. Scoring was done by two independent judges, with an average percentage of agreement of 79 for individual categories.

With respect to the conception of external reality, a positive view was inferred from any reference to a pleasant or helpful environment, indications of curiosity about the world, a sense of predictability about it, a preference for reality rather than myth, and pleasure in the exercise of responsibility.

TABLE 4

PROJECTIVE QUESTIONS AND CRITICAL RESPONSE CATEGORIES

Item and Variable	Code Category
Questions on the conception of external reality	
1. What things would you say help a person most to get ahead these days?	
a. uncontrollable world	Exclusive mention of luck, chance, family connections
b. controllable world	Any mention of hard work, ambition, getting along with people
2. Children should have no worries, they should be protected from troubles as much as possible	
a. oppressive world	Agreement with statement
Reasons for agreement or disagreement	
b. oppressive world	The world is too difficult a place for children, (or) the world is difficult and they might as well learn that it is
3. Do you think it's a good idea or a bad idea to tell children that babies are brought by the stork?	
a. preference for myth	Good idea to tell children the stork story
4. What do you think makes some people worry a lot when other people hardly ever worry?	
a. oppressive world, uncontrollable	Too many responsibilities, poverty, family troubles, death of loved ones, other disasters
5. What would you say was the nicest thing about being a child?	
a. responsibility oppressive	Lack of responsibility
6. Children should be taught to fight their own battles as early as possible.	
a. harsh social world	Strong agreement because world is viewed as harsh, a jungle
Questions on the acceptance/rejection of normal Impulses	
7. (Suppose) a person has an awful lot of worries and they're really getting him down. What do you think he should do about it?	
a. impulse denial	Pull self together; exercise self-control, stop thinking about it, rise above it, occupy mind with other things, etc.

TABLE 4—*continued*

Item and Variable	Code Category
8. Some people feel that it's good for everyone to get really angry now and then and "blow off steam." Others say its childish to let yourself feel so angry. How do you feel about this?	
a. impulse denial	It's childish to let yourself feel so angry
9. Parents should plan out the child's program each day so they're sure he is learning the right things	
a. suspicion of untrained impulses	Agreement with statement
Reasons for agreement of disagreement *b.* suspicion of impulses	Child is basically bad, has to be directed into good, (or) no planning can prevent the child from learning bad things—it's in his nature
10. Parents should make children try things that are a little beyond them in order to make them ambitious	
a. suspicion of untrained impulses	Agreement with statement
Reasons for agreement *b.* suspicion of impulses	Child is by nature lazy, will not strive if not pushed
11. Reasons for telling children that babies are brought by the stork (question 3)	
a. rejection of impulses	Sex is bad, immoral, evil
12. If a daughter of yours were getting married, what things would you want to know about the man she was marrying?	
a. suspicion of impulses	Whether he was a drinker, philanderer, criminal

Denial of impulses was inferred from any indication of suspicion about untrained impulses or rejection of normal expression of emotionality.

B. Analysis

In the results presented below, tests involving responses to the projective questions were made on 314 cases. The two remaining respondents did not answer one or another of the questions included in the competence index. The chi-square technique was employed for testing differences, and product-moment correlation coefficients were computed.

Predictions about the story responses to the TAT were based on the subsample of 76 cases in which this measure was administered. These 76 cases were separated into three groups on the competence index (low and very low were combined). In the tables referring to story responses, percentages are based on the total number of stories rather than the number of respondents. Since such a procedure may have an inflative effect on differences when there is a positive intraclass correlation within clusters of responses given by a single respondent, it was necessary to use a formula for the variance of a proportion which accommodates for the effect of clustering.[7] Since the use of such a formula is permissible

---

[7] The following formula, conventionally used in survey analysis of clustered responses (6, p. 208), was used:

$$S_p^2 = \frac{1}{m(m-1)A^2}\sum_{1}^{m} x_i^2 - \frac{1}{m}\left(\sum_{1}^{m} x_i\right)^2$$

where:

- m equals the number of clusters (i.e., individuals);
- A equals the cluster size (i.e., number of story-responses per individual);
- X_i equals the total score of the *i*th individual.

TABLE 5

SCORING CATEGORIES FOR THE TAT STORIES

Category scheme and variable	Category
I. Conception of External Reality	
1. Nature of response	
a. curiosity about a new task	Responds with an imaginative story
b. apprehension about a new task	Responds with card-description or some form of rejection of the task
2. Setting of the story	
a. oppressive environment	Stories in which the theme is one of environmental oppression, in contrast to those in which the theme concerns personal or interpersonal conflict
3. Need of central character	
a. curiosity about environment, movement outward	Need for achievement, independence
b. oppressive environment	Need for relief, dependency, succorance, or escape
4. Nature of environment	
a. oppressive environment	Environmental blocks obstruct the movement of the central character
5. Action of central character	
a. movement outward, utilization of skills	Goal-directed action by central character
6. Outcome of the story	
a. curiosity about environment	Completion of story
b. noncurious about environment	Unresolved ending
c. oppressive environment	Failure ending
d. utilization of skills	Success ending accompanied by instrumental activity
II. Acceptance/rejection of normal impulses	
7. Feeling expressed by central character	
a. impulse denial	No mention of feelings of central character

only in the case of a fourfold table, tests were made between the extreme competence groups. The tests are presented in terms of *t* values, since the test of proportions is the same as chi square in the case of a fourfold table. All three competence categories were used in computing correlation coefficients.

C. RESULTS AND DISCUSSION

In Tables 6 and 7 we present findings from the projective questions and the TAT stories that refer to the respondents' conceptions of the outside world. We find that individuals who score low on the personal competence index consistently give more answers which reflect a picture of an oppressive, uncontrollable external reality. Compared to high scorers, they more often attribute success (item 1) exclusively to chance, luck, or other factors beyond individual control. More often than high scorers they take a somewhat overprotective position on child training, and give reasons which indicate that the world is a difficult place (item 2). They prefer myth to reality (item 3) more often than individuals who feel competent. On items 4 and 5 which indicate, respectively, attitude toward responsibility and the attribution of worry to external factors, the differences which occur are in the predicted direction but do not meet our criterion of significance.

The single item in this area which does not even order the competence groups in a manner consistent with our prediction is one (item 6) in which reality is seen not merely as oppressive but actually hostile. This response is similar to the "world as a jungle" conception of the authoritarian personality, and is the only one of our items which clearly implies projection of repressed hostility.

In analysis of the TAT protocols we find again evidence that respondents who do not feel competent have less optimistic expectations about the world around them. They manifest less curiosity about a new stimulus and are less active in approach to the environment. They reject the story-telling task more often, and the responses they do give are more often bare descriptions of the cards or minimal stories, less often imaginative stories. Themes of stories given by people who score low on competence more often center on a man's struggle against external forces, more often picture the environment as oppressive rather than helpful, and more often are unresolved or have failure resolutions.

Those people who feel competent are more likely to reveal in their stories and answers an active interest in the environment, and anticipation of a pleasant, manageable world. They portray man in conflict with himself or others, striving to achieve and become independent, and reaching his goals through directed effort. Responsibility is central to this view, and is assumed with a sense of gratification.

Of twenty tests (8 in Table 6 and 12 in Table 7) made of the hypothesis relating the sense of competence to the nature of the individual's conception of external reality, eleven yielded results significant beyond the 5 per cent level of confidence. Eight of the remaining nine tests showed differences in the predicted direction, but did not meet this criterion. Four of the eight have p values between .05 and .10, and four have p

TABLE 6

PROPORTIONS OF COMPETENCE GROUPS GIVING CRITICAL RESPONSES TO PROJECTIVE QUESTIONS

Item and critical category	Competence				χ^2	p	r
	Very High (N=116)	High (N=102)	Low (N=60)	Very Low (N=36)	(df=3)	level	
Conception of external reality							
Item 1							
a. uncontrollable world[a]	.18	.25	.33	.37	6.62	.10–.05	.20
b. controllable world[a]	.40	.36	.24	.19	14.00	.01	.25
Item 2							
a. oppressive world	.50	.64	.59	.70	6.77	.10–.05	.18
b. oppressive world	.28	.40	.37	.55	9.97	.02–.01	.23
Item 3							
a. preference for myth	.29	.33	.37	.58	9.78	.05–.02	.21
Item 4							
a. oppressive world, uncontrollable	.12	.22	.29	.31	4.87	.20–.10	.15
Item 5							
a. oppressive world	.60	.68	.72	.75	4.69	.20–.10	.16
Item 6							
a. harsh social world	.18	.17	.20	.17	.183	.99–.98	.01

[a] This question was asked of only 236 respondents. The numbers in the competence categories should read: Very High 81, High 83, Low 41, Very Low 31.

TABLE 7

PROPORTIONS OF COMPETENCE GROUPS GIVING CRITICAL RESPONSES TO TAT

Category	Competence					
	Very High 1 ($N=99$)	High 2 ($N=51$)	Low and Very Low 3–4 ($N=81$)	t (High vs. Low)	p value	r
I. Conception of external reality						
1. Nature of response						
a. imaginative story	.55	.41	.35	2.1377	.02	.25
b. rejection of task	.18	.23	.25	.955	.20–.10	–.14
2. Setting of story						
a. environmental oppression	.39	.43	.52	1.572	.10–.05	–.19
b. personal and/or interpersonal conflict	.56	.49	.42	2.174	.02	.21
3. Need of central character						
a. independence or achievement	.66	.55	.46	2.486	.01	.26
b. dependency, succorance relief	.31	.39	.51	2.475	.01	–.24
4. Nature of environment						
a. oppressive obstructive	.26	.41	.46	2.163	.02	–.22
b. pleasant, helpful	.12	.06	.07	1.099	.20–.10	.13
5. Action of main character						
a. goal directed	.46	.29	.19	4.219	.01	.36
6. Outcome						
a. success	.52	.31	.37	1.783	.05	.20
b. unresolved	.26	.41	.37	1.296	.10–.05	–.17
c. success with instrumental activity[a]	.70	.56	.41	3.8173	.01	.33

[a] In this case, the number on which proportions are based consists of the number of successful stories told. The numbers for the three competence groups were: Very High 48, High 16, Low and Very Low 32. The formula for computing the variance of proportions was in this case one designed for use with unequal sized clusters (6).

values between .10 and .20. The consistency of these differences may be regarded as indicating a trend in the predicted direction.

In relating the competence variable to responses to *internal* reality, on the other hand, we found no significant results in the nine tests made. Our low competence group seems no less aware or accepting of normal impulses and emotionality than those scoring high on the index of competence. They do not mention feeling states less frequently in their TAT stories (the p value falls between .30 and .40). They are no more likely to reject expression of impulses and feelings (the p values for differences on items 7 and 8 in Table 4 fall between .30 and .50; for item 11 the p value is between .90 and .95). And they show no more suspicion of untrained or unsupervised impulsivity (items 9, 10, and 12 of Table 4; p values are between .50 and .70, between .10 and .20, and between .70 and .80 respectively). In the one case where a difference approached significance (item 10, p falls between .10 and .20), the high competence group was somewhat more resistive to normal impulses than were those who scored low on competence.

We conclude from these results that the conscious feeling of competence rests on the nature of the individual's con-

ception of the outside world, but that at least as measured in our study, the person's mode of response to the internal environment does not contribute significantly to the competence variable.

The following tentative interpretation of these results may be offered. The sense of personal competence—a variable we have found useful in predicting a person's orientation toward public affairs—implies a binding commitment to external reality. Assuming that no reality is perfectly gratifying—that everyone's life situation involves some degree of frustration—the person who says "I can to a significant extent determine the conditions of my life," is committing himself to attend and to affect the world he lives in. The opposed position ("I am not in control of what happens") provides a satisfactory escape from such a commitment. If one has no control over the situation in any case, there is also no need to maintain a high level of interest in the world around.

This does not mean, of course, that some individuals who feel they have no control will not at the same time be active and interested in the external world. There are obviously other factors which contribute to the individual's involvement in events outside himself. Social pressure, conformity to primary group standards, and the "play motive" (8) are among the many forces that may promote an interest in public affairs quite apart from any conception the individual has concerning his power to affect the world. Since we have no reason to assume that these other factors are differentially distributed in the competence groupings, we have assumed in this discussion that the reader will supply the phrase "other things being equal." In the present discussion we are interested in why the sense of competence should influence a person's attentiveness with respect to public affairs. The interpretation of the sense of personal competence as implying a commitment to attend seems to clarify this relationship.

To make a strong commitment to events outside oneself would seem to require a view of the world as at least potentially manipulable and gratifying. Conversely, it is reasonable to assume that a person who takes a position legitimizing withdrawal from reality would also hold a more awesome picture of the outside world.

But the question of the defense foundation of these two views does not appear critical. One may conceive an unchangeable environment—or changeable but beyond any measure of personal control—either because one has had consistent experiences with a difficult environment, or because one projects an unmanageable (and denied) set of impulses outward onto relatively innocuous surroundings. Whatever its origin, the anticipation of satisfaction or oppression from the world appears to enter into the determination of the sense of personal competence and the level of activity with which one confronts the world.

Our results confirm Mussen and Wyszinski's (7) finding that the apathetic individual has a "generally external orientation" and shows marked passivity toward the environment, but they do not support the interpretation of these authors that this passivity is a defense against hostile impulses that have been externalized. While Mussen and Wyszinski speculate that the difference between the apathetic and ethnocentric individuals lies in the nature of their response to a world invested with hostility, our data suggest that the two types also differ in the nature and derivation of the world views they hold.

The withdrawal of our low-competence group does not seem to stem from a view of a hostile world. Rather, the evidence indicates that the crucial aspect of their world view is the element of intractability. Reality is oppressive and uncontrollable, but not necessarily permeated by personal hostility that has been projected because it is unacceptable as part of the self. The results of questions about impulses do not reveal a base of denial-projection underlying their sense of a burdensome reality.

IV. SUMMARY AND CONCLUSIONS

The study from which this article derives was a sample survey of labor force members in the city of Detroit. The results of the study can be summarized under two major findings:

1. The sense of effectiveness in public affairs has been shown to be moderately but consistently related to a feeling of competence in a variety of areas in the more immediate environment of personal affairs. The questions used to measure attitudes toward the narrower personal sphere varied both in degree of specificity and in content (e.g., financial and occupational affairs, and personal planning). The fact that effectiveness was related to the sense of personal competence in a variety of areas we interpreted as an indication that this attitude toward one's role in public affairs is in many people a reflection of a more general orientation toward the world.

2. Competence was analyzed in relation to two aspects of personality—the nature of the person's conception of the outside world, and the acceptance or rejection of normal impulses. These variables were measured by a series of projective questions—primarily in the content framework of child-rearing philosophies—and by a three-card TAT measure administered to a subsample of the study population. The competence variable showed low but consistent relations to the projective measures of the person's attitude toward external reality, but did not relate to the acceptance or rejection of normal impulses.

The primary conclusion we draw from the total study concerns the role of ego processes as mediating variables in the determination of social attitudes. In the study of attitudes toward social issues, a question which arises repeatedly is: To what extent do such attitudes reflect a rational response based on the real event in combination with some notion of self-interest; and to what extent is there an incursion on this rational process by more personal aspects of the individual's psychological integration—stemming from a unique history of experiences, and not necessarily conscious or realistic?

This question is of interest both because of the general scientific goal of understanding attitude determination, and also because of its implications for the practical problem of producing change. Educational techniques are designed to affect attitudes by changing the reality they reflect. To the extent that social attitudes are nonrationally determined, we must modify either our goals or methods in planning efforts of change.

What our analysis reveals, we believe,

is that to a statistically significant though small degree, personality variables do contribute to the person's orientation toward public affairs. We have seen that the sense of effectiveness is consistently related to a more general personal orientation—whether measured in questions dealing with everyday planning or in projective devices designed to elicit less conscious attitudes toward the world. The fact that the sense of effectiveness has been shown to be in many people a part of a general way of looking at the world—that governmental affairs become a vehicle for the expression of a personal orientation—suggests that whatever external factors enter as determinants of this attitude are at least to some extent mediated through a screen of stable personality characteristics.

Assessing the relative contribution of rational and nonrational processes in the determination of this attitude requires further research and more refined analyses. The finding that the sense of competence relates to the view of external reality but not to the use of denial as a defense against impulses requires further confirmation, but is provocative in its discrepancy from the bulk of work on the relation of social attitudes to personality variables. Additional research might lead to a clarification of the role of defenses in the determination of social attitudes.

REFERENCES

1. Adorno, T. W., Frenkel-Brunswick, Else, Levinson, D. J., & Sanford, R. N. *The authoritarian personality.* New York: Harper, 1950.
2. Buchanan, W. Mass communication in reverse: the Unesco international survey. *Int. Soc. Sci. Bull.,* 1953, 5, 577-584.
3. Campbell, A., Gurin, G. & Miller, W. E. *The voter decides.* Evanston, Ill.: Row, Peterson, 1954.
4. Douvan, Elizabeth, & Walker, A. *The impact of atomic energy on society: an investigation of attitudes and their determinants.* Ann Arbor, Mich.: Survey Research Center, 1953. (Mimeographed.)
5. Gouldner, A. W. (Ed.). *Studies in leadership.* New York: Harper, 1950.
6. Kish, L. Selection of the sample. In L. Festinger & D. Katz (Ed.). *Research methods in the behavioral sciences.* New York: Dryden, 1953. Pp. 175-239.
7. Mussen, P. H., & Wyszinski, Anne B. Personality and political participation. *Hum. Relat.,* 1952, 5, 64-82.
8. Riesman, D. *The lonely crowd.* New Haven: Yale Univer. Press, 1950.
9. Riesman, D., & Glazer, N. Criteria for political apathy. In A. W. Gouldner (Ed.). *Studies in leadership.* New York: Harper, 1950.
10. Wert, J. E., Neidt, C. O., & Ahmann, J. S. *Statistical methods in educational and psychological research.* New York: Appleton-Century-Crofts, 1954.

(Accepted for publication December 17, 1955)

ABILITAT ON DIA
CALCUTT
700068